MENTAL SYMBOLS

STUDIES IN COGNITIVE SYSTEMS

VOLUME 19

EDITOR

James H. Fetzer, *University of Minnesota, Duluth*

ADVISORY EDITORIAL BOARD

Fred Dretske, *Stanford University*

Charles E.M. Dunlop, *University of Michigan, Flint*

Ellery Eells, *University of Wisconsin, Madison*

Alick Elithorn, *Royal Free Hospital, London*

Jerry Fodor, *Rutgers University*

Alvin Goldman, *University of Arizona*

Jaakko Hintikka, *Boston University*

Frank Keil, *Cornell University*

William Rapaport, *State University of New York at Buffalo*

Barry Richards, *Imperial College, London*

Stephen Stich, *Rutgers University*

Lucia Vaina, *Boston University*

Terry Winograd, *Stanford University*

The titles published in this series are listed at the end of this volume.

MENTAL SYMBOLS

A Defence of the Classical Theory of Mind

by

PETER NOVAK

SPRINGER SCIENCE+BUSINESS MEDIA, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-94-010-6374-6 ISBN 978-94-011-5632-5 (eBook)
DOI 10.1007/978-94-011-5632-5

Printed on acid-free paper

All Rights Reserved
© 1997 Springer Science+Business Media Dordrecht
Originally published by Kluwer Academic Publishers in 1997
Softcover reprint of the hardcover 1st edition 1997
No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any information storage and
retrieval system, without written permission from the copyright owner.

SERIES PREFACE

This series will include monographs and collections of studies devoted to the investigation and exploration of knowledge, information, and data-processing systems of all kinds, no matter whether human, (other) animal, or machine. Its scope is intended to span the full range of interests from classical problems in the philosophy of mind and philosophical psychology through issues in cognitive psychology and sociobiology (regarding the mental abilities of other species) to ideas related to artificial intelligence and computer science. While primary emphasis will be placed upon theoretical, conceptual, and epistemological aspects of these problems and domains, empirical, experimental, and methodological studies will also appear from time to time.

In this unusual volume, Peter Novak launches an unrelenting attack upon recent analytic approaches toward understanding the nature of the mind, including the Conservative, the Radical, and the Middlebrow positions that he identifies with Fodor, Quine, and Putnam, respectively. In their place, Novak advocates (what he calls) The Classical Theory of Mind, which builds upon an ontology of simple and complex concepts, generative mechanisms for producing sentences and cognitive mechanisms that govern both cognitive and emotional operations, combined with a Kantian epistemology that separates what can and cannot be accessible to the mind. This study, which the author intends as a critique of the Analytic Movement in philosophy, may infuriate as many as it fascinates, but it will not be easy to ignore.

James H. Fetzer

Table of Contents

Introduction
A Fly in a Bottle xi

Chapter 1
Conservative Rationalism I 1

1.1 Common-sense Psychology 1
1.1.1 The language of thought 2
1.1.2 The computer analogy 3
1.1.3 The referential theory of meaning 5
1.2 The Supervenience Chain 8
1.2.1 Syntactically identical symbol-tokens with different extensions 9
1.2.1.1 Narrow and broad meaning 10
1.2.2 Complex symbols under-determining extensions 12
1.2.3 Syntactically distinct complex symbols with identical extensions 14
1.3 The Nomic Theory of Reference 16
1.3.1 The four-phase nomic theory of reference 17
1.3.1.1 The physical phase 18
1.3.1.2 The psychophysical phase 19
1.3.1.3 The psychological phase 19
1.3.1.4 The sociological phase 19
1.3.2 An error in an account of error? 20
1.4 *Casus Belli* 21
1.5 Nomenclature 22

Chapter 2
The Idea as World and Will
A Belief in a Contribution of Environment to Meaning and a Division of Semantic Labour 25

2.1 The Ambiguous Meaning of "Meaning" 25
2.1.1 Scientists as semantic experts 25

2.1.2 A Middlebrow compromise between extensionalism and mentalism 26
2.1.3 Pragmatic realism 29
2.2 The Will to Linguistic Power 32
2.2.1 Semantic oligarchy: scientists as unrepresentative leaders 32
2.2.2 A Westminster system? Extensions as Lords, stereotypes as Commons 35
2.2.3 Колхоз semantics, or meaning holism 35
2.2.4 Semantic tyranny: Merlin knows and rules by meaning alone 36

Chapter 3
Sentence-based Semantics

Early Steps toward Semantic Holism 39

3.1 Motivation for Semantic Holism 39
3.2 Frege on Term-based Semantics 41
3.2.1 Imagistic mentalism: variability in *vorstellungen* 43
3.2.2 Number-referring terms in axiomatic proofs 45
3.3 Russell on Term-based Semantics 47
3.3.1 Meaningful terms referring to non-existent objects 50
3.3.2 The sentence-based semantics of denoting phrases 50
3.4 Carnap's Semantic Verificationism 52
3.4.1 Sentences as fundamental bearers of meaning 55
3.4.2 Meaning as method of verification 55
3.4.3 Contingent verification as confirmation of indefinitely many predictions 56

Chapter 4

Radical Empiricism I 59

4.1 Meaning and Non-existent Entities 59
4.2 The Separation of Meaning and Reference 60
4.2.1 The nominalistic extensional theory of meaning 64
4.2.2 The realistic extensional theory of meaning 65
4.2.2.1 The conceptual-scheme relativity of what there is 65
4.2.2.2 The explanatory power and simplicity of a conceptual scheme 67
4.2.2.3 The actual use of language, and actual practices of positing universals 68

4.3 Mental and Mind-Independent Semantic Universals . . . 69
4.3.1 The bad metaphysics charge 70
4.3.2 The *virtus dormitiva* charge 72
4.3.3 The undefinability charge 73

Chapter 5

Radical Empiricism II 77

5.1 Two Major Routes to Semantic Holism 77
5.2 The Route from Verificationism to Holism 77
5.2.1 Synthetic statements 79
5.2.1.1 Phenomenalism in natural science 80
5.2.1.2 Universal epistemic revisability 80
5.2.1.3 The inference from confirmation holism to semantic holism 83
5.2.2 Analytic statements 85
5.3 The Route from Behaviourism to Holism 87
5.3.1 The argument from collateral information 90
5.3.2 The argument from inscrutability of reference 92
5.3.3 Four refutations of behavioural holism 93
5.3.3.1 The language-learning argument 94
5.3.3.2 The premiss that meaning is what a sentence shares with its translation 95
5.3.3.3 The argument from collateral information 96
5.3.3.4 The argument from inscrutability of reference 97
5.4 Remarks on Middlebrow Pragmatism 100
5.4.1 Meaning holism 100
5.4.2 The coherence theory of truth 101
5.4.3 The satisfaction theory of reference 103
5.4.4 Holism's wake 105

Chapter 6

Conservative Rationalism II 107

6.1 The Old Sorcerer's Supervenience Chain 107
6.2 Misrepresentation and Asymmetric Dependence 110
6.3 The Sociological Phase of Reference 116

Chapter 7
The Classical Theory of Mind I 125

7.1 Five Ways of Defining Logical Modality 125
7.2 The Classical Theory of Mind 127
7.2.1 There is a *finite* basis of semantically *simple ideas* (taken as *types* of symbol), or simple terms of the mind's representational code 127
7.2.2 There is a *generative mechanism* comprising operations for the production of infinitely many *complex ideas* from the empirical basis 128
7.2.3 There is a generative mechanism comprising operations for the production of infinitely many *propositions* from the stock of simple and complex ideas 129
7.2.4 There are finitely many basic *psychological operations* on propositions 130
7.2.5 There is a generative mechanism for the production of *complex psychological operations* on propositions . . 131
7.3 The Classical Theory of Representation 132
7.3.1 Basic ideas 133
7.3.2 Complex ideas, and the *a priori* ideas generating them . 133
7.3.2.1 *The simple* a priori *ideas* $\neg$ *and* $\wedge$ 133
7.3.2.2 *The complex* a priori *ideas* $\vee$, $\supset$, *etc.* 134
7.3.2.3 *The simple* a priori *predicate* $=$ 134
7.3.2.4 *Complex empirical ideas* 134
7.3.3 Propositions, and the *a priori* ideas generating them . . 135
7.3.3.1 *The simple* a priori *idea* Σ 135
7.3.3.2 *The complex* a priori *idea* A 136
7.3.3.3 *Propositions* 136
7.3.3.3.1 *Identity propositions* 136
7.3.3.3.2 *Quantified propositions* 136
7.3.3.3.3 *Compound propositions* 136
7.4 Aspects of the Classical Theory of Knowledge 137
7.4.1 Locke on epistemic evaluation 137
7.4.2 Locke's prevision of a new 'Logick and Critick' 138
7.4.3 Metalogical principles and definitions 141
7.5 Modal Properties in the Model Code 143
7.5.1 Analytic evaluation in the model of CTM 143
7.5.2 The trifling propositions $a=a$ and $(\mathrm{A}x)(Fx \supset Fx)$ 149
7.5.3 Comments on innate and *a priori* knowledge 151
7.6 The Complex Ideas of Implication 152
7.6.1 Assuming and implying in a true conditional 152
7.6.1.1 Assuming a logical falsehood 152
7.6.1.2 Implying a logical falsehood 153

7.6.1.3 Assuming a logical truth 153
7.6.1.4 Implying a logical truth 154
7.6.2 The complex ideas $\rightarrow$, $\rightarrowtail$, $\Rightarrow$, and $\rightharpoonup$ 155
7.6.2.1 The complex *a priori* ideas of necessary implication . . 155
7.6.2.2 The complex *a posteriori* ideas of contingent implication 158
7.6.3 Comments concerning the normativity of complex ideas 160
7.7 The Complex *A Priori* Idea of Valid Deductive Inference 160
7.7.1 The complex *a priori* idea $\vdash$ 161
7.7.2 Some of Quine's objections against logical modality . . . 162
7.8 Remarks on CTM and Analytic Philosophy 163

Chapter 8
The Classical Theory of Mind II 167

8.1 The Ontology and Architecture of CTM 167
8.2 Learning, Memory, and Association in *Aplysia* 170
8.3 Learning, Memory, and Association in Vertebrates . . . 175
8.4 The Genetic Code, the Mental Code 179
8.4.1 The Classical Theory of Mind 180
8.4.2 A conjecture on the nature of mental symbols 181
8.4.3 Cognitive architecture 181
8.4.3.1 The vertical dimension 183
8.4.3.2 The horizontal dimension 185
8.4.4 Concept-acquisition and memory 187
8.4.5 Recall and simple associations 187
8.4.6 Rational processes 188
8.5 History, Histology, and the Molecular Level of Analysis 189

Chapter 9
The Classical Theory of Mind III 195

9.1 Toward an Integrated Account of Mind 195
9.2 The Symbolic System of a Deep-layer, Long-term-store Psychic Cell 195
9.2.1 Syntax 196
9.2.2 Semantics 199
9.2.3 Epistemology: cognitive processes in the psychic cell . . 201
9.2.3.1 *A priori* knowledge 202

9.2.3.1.1 Analysis 202
9.2.3.1.1.1 Implication 203
9.2.3.1.1.2 Inference 204
9.2.3.1.2 Synthesis 204
9.2.3.2 *A posteriori* knowledge 210
9.2.3.2.1 Observational, *a posteriori ex terminis* synthesis 211
9.2.3.2.2 Holistic *a posteriori* synthesis 212
9.3 The Mind as a System of Psychic Cells 213
9.3.1 The acquisition and retention of ideas 214
9.3.2 The recollection and remembrance of ideas 215
9.3.3 Associative processes 216
9.3.4 Rational processes 216
9.4 The Mind and its External Affairs 217
9.4.1 Adaptation in the natural and social environment 217
9.4.2 Mind and language 218
9.4.2.1 Words standing for simple ideas 218
9.4.2.2 Words standing for complex ideas 219
9.4.2.3 Statements standing for propositions 220
9.4.3 Levels of intellection and education 220

Chapter 10
The Tale of Russell's Paradox 225

10.1 The Misty Origins of Analytic Philosophy 225
10.2 Russell's Sophism in the Classical Theory of Mind . . . 230
10.2.1 Levels of being 232
10.2.1.1 Metaphysical reality 232
10.2.1.2 Physical reality 233
10.2.1.3 Psychological reality 234
10.2.1.4 Social reality 235
10.2.2 Preliminaries: the smallest and the greatest set 236
10.2.3 The truth-condition of the sophism 238
10.3 The Well-Formedness of the Mind 241

Epilogue 243
Bibliography 247
Index 259

Introduction

A Fly in a Bottle

Analytic Philosophy of mind and meaning has operated throughout this century much like a two party political democracy, a Conservative Party at loggerheads with a Radical Party, and an obligatory minor Middlebrow Party straddling the major positions and holding the balance of power. Currently, among the key features of the Conservative platform are the following:

(i) The mind is a system of operations, identifiable with believing, desiring, *etc.*, on tokens of the sentences of a representational mental code, generated from tokens of terms of the code, or concepts.

(ii) The mind's overall organisation, or architecture, is analogous to the classical computational architecture of the von Neumann machine, with a central processor in which terms and sentences of the representational code are tokened, and in which cognitive processes take place serially. (The Conservatives pride themselves on having here the only workable account of rational cognitive processes available.)

(iii) The meaning of a public symbol (word, statement, *etc.*) derives from that of the corresponding mental symbol.

(iv) The referential theory of meaning, together with a nomic or causal theory of reference, holds for mental terms, and the truth-conditional theory holds for mental sentences. In general, the principle of extension-meaning supervenience is accepted, in that all tokens of a type-identical mental symbol are taken to refer to identical extensions or truth-conditions. (The Radicals usually twist this by saying that the Conservatives believe the extensional theory of meaning, that the meaning of a symbol *is* its extension.)

(v) Conceptual definability does not hold, and neither does any definitional account of reference (by sufficient and necessary conditions for the membership in a term's extension). Consequently, most or all terms of the mental code are supposed to be semantically simple, innate, and universal for the human mind; an extreme version of rationalism is accepted.

(vi) The traditional distinction between *a priori* and *a posteriori* knowledge is taken seriously. *A priori* knowledge is assumed to be knowledge independent of experience, whilst *a posteriori* knowledge, knowledge

dependent on experience. There are propositions known *discretely*, either *a priori* or solely by observation; global epistemic holism is rejected.

(vii) The traditional distinction between analytic and synthetic propositions is also upheld. Analytic propositions are supposed to be those true by virtue of meaning, synthetic those true by virtue of empirical fact; analyticity and syntheticity are regarded as mutually exclusive, so that an analytically true proposition cannot be synthetically true, and conversely. Truth is seen as an objective, absolute semantic value, and as contradictory to falsehood.

(viii) Logical modalities, entailment, validity, and nomological and other modalities are accounted for in terms of possible worlds, though not without misgivings. Possible worlds are treated with suspicion, but put up with for pragmatic reasons.

(ix) Ontological materialism is adopted, but not nominalism. Natural kinds abound: mental states and processes are of a natural kind; and semantic properties are also natural kinds.

So much for sketching the Conservative platform. There is a common ground underlying the Conservative and the Radical agendas, but on the surface the Radicals are very different:

(a) Mental states and processes are not of a natural kind, and as such do not exist; there are no mental symbols (concepts, ideas), and no such things as minds, human or other.

(b) Between the ears, there is a nervous system which functions as a dispositional mechanism, mapping external inputs into determinate behavioural outputs. The dispositional mechanism varies from individual to individual and, for each individual, from time to time, depending on its initial conditions and its subsequent history of inputs. The organisation, or architecture, of the mechanism is not analogous to the classical computational architecture: there are no symbols involved, and no central symbolic processor. Instead, the mechanism is a network of inter-connected nodes which more or less transmit activity between inputs and outputs. (The Radicals pride themselves on having here an account of dispositional processes which matches what is known of the structure and function of the brain.)

(c) The meaning of a public symbol (word, statement, *etc.*) is not derived from that of any corresponding mental symbol; public symbols are the only symbols there are, and are the only bearers of meaning.

(d) The principle of extension-meaning supervenience is rejected, in that different tokens of a word or statement need not have identical extensions or truth-conditions; meaning does not determine extension, but under-determines extension. Still, the referential theory of meaning, and nomic or causal theory of reference, are accepted, provided we now think of reference not as a relation of determining the extension of a word, but as a relation of the use and application of the word with respect to resemblance-bound stimulatory conditions, either factual or counter-factual.

(The Conservatives regard this dissipation of reference and meaning as an especially vicious Radical subversion.)

(e) Semantic definability is denied, in accord with the Conservatives, not only because there are no sufficient and necessary conditions for the membership in the extension of a term, but mainly because the principle of extension-meaning supervenience does not hold: the term does not have a determinate extension. Further, there is no basis of semantically simple terms universal for human languages; semantic innateness and universality are altogether rejected, and an extreme version of empiricism is adopted.

(f) The traditional distinction between *a priori* and *a posteriori* knowledge is rejected, but the notions are shared with the Conservatives: *a priori* knowledge is regarded as knowledge independent of experience, *a posteriori* as knowledge dependent on experience. No propositions are known discretely, either *a priori* or by observation; global epistemic holism is accepted.

(g) The traditional distinction between analytic and synthetic propositions is also dismissed; but again, the notions are shared: analytic propositions, if there were any, would be true by virtue of meaning, synthetic true by virtue of empirical facts (no propositions being both analytic and synthetic). Truth is regarded as a semantic value, though not as absolute and objective, nor as unrevisably opposed to falsehood; it is a matter of coherence in a conceptual scheme, rather than of correspondence to empirical facts or of pure semantic determination.

(h) Logical necessities, entailments, and valid arguments are but conventions we would prefer not to revise, but would revise under duress; and once revised, they would have whatever modality they might be said to have in the new behavioural-*cum*-linguistic scheme. Possible worlds are viewed as fictions, useful — along with other fictions — in formalising and regimenting some areas of linguistic behaviour.

(i) Ontological materialism is shared with the Conservatives, but nominalism is pushed farther. Natural kinds and some doctrinally sound abstractions may be allowed; but mental states and processes are *not* of a natural kind, and semantic properties are anathema.

Much of this century, the Radical Party has been in office, not because the electorate found its programme and performance satisfactory, but because the Conservatives rendered themselves ineffective by their extreme rationalism, and that *lapis philosophorum*, the principle of extension-meaning supervenience. Hence, perhaps, the Middlebrow Party has arisen, running on the policy that the truth is somewhere in the middle, between the right and the left, or if need be, both on the right and on the left. The Middlebrow view belongs to those who know the experience of 'being torn':

(1) The mind is not something, but it is not nothing either. Mental states are not of a natural kind, and as such are not scientifically

individuable; still, one way or another, they emerge from behavioural patterns and dispositions.

(2) The mechanism mediating between stimulatory inputs and behavioural outputs is not organised like a classical von Neumann machine, but as a connectionist neural network which may vary from individual to individual and from time to time, depending on its initial conditions and its history of inputs. Yet the network is not merely a dispositional mechanism; concepts or ideas, beliefs, desires, intentions, *etc.*, emerge from the behavioural dispositions and patterns, and emergent rational processes occur.

(3) Public symbols are not the only symbols there are; mental representations can be admitted, though not as constituting a discrete mental code, and not as the primary bearers of meaning. The meaning of a public symbol comes rather from its overt use and application; but it is also fixed by the nature of the environment to which the symbol is applied.

(4) The principle of extension-meaning supervenience is repudiated, and upheld. Meaning does not determine extension, but under-determines it; different tokens of a public word or statement need not have identical extensions or truth-conditions. The environment, however, contributes to meaning, and is the final arbiter on the semantic identity not only of natural-kind words, but also of words referring to artifacts. Different speakers have different *epistemic* access to the environment, and so to *meaning*; there is a division of semantic labour: knowing the environment *is* knowing meanings. Further, reference is explained both causally — in terms of indexically fixed aspects of the environment and the division of semantic labour — *and* holistically, in terms of verbal use and trivial satisfaction. 'What does "rabbit" refer to? Why, to rabbits! What does "extraterrestrial" refer to? To extraterrestrials!'

(5) Semantic definability and any definitional account of reference are denied: there is no basis of semantically simple terms universal for all human representational systems, and there are no sufficient and necessary conditions determining the membership in the extension of a term. Accordingly, semantic nativism is rejected and a version of empiricism accepted, not without a rationalist flavour.

(6) The traditional distinction between *a priori* and *a posteriori* knowledge is said to be but a convention, and global epistemic holism is approved; still, the convention is not dismissed as arbitrary: all knowledge, *a priori* or *a posteriori*, is conceptual-scheme relative, yet it is nevertheless objective.

(7) The traditional distinction between analytic and synthetic propositions is also an objective convention; whether a proposition is analytically or synthetically true depends on how it is used (is it used correctly come what may, or is it used correctly of such and such a situation?). Truth is explained by reference to correct use, and correct use by reference to warranted use (without being 'cute'). Truth is not absolute;

it is a matter of coherence in a conceptual scheme; yet it is still objective and independent of what most people believe.

(8) Logical necessities, entailments, valid arguments, *etc.*, are likewise objective verbal conventions, only extra-coherent and super-conventional. Ersatz possible worlds might be acceptable, conceptual-scheme permitting.

(9) What there is, is relative to what conceptual scheme is in place; relativism is rejected, though. It is not that each conceptual scheme cuts up the world, the reality, in its peculiar way, some being more true of the world than others; rather, there would be no world, no reality, were it not for a conceptual scheme — a public language — at work. For example, one would not exist unless someone had a public language to say one does exist, and said so; one would be nothing at all without public recognition. (Here, especially, the strength of the Middlebrow view comes to the fore.)

Some readers may have discerned the voices of the leaders of the three parties: Jerry Fodor as leader of the Conservative Party, Willard Quine as leader of the Radical Party, and Hilary Putnam as leader of the Middlebrows. On its polemical side, the purpose of this book is to argue against these Analytic trends, to the general conclusion that Analytic Philosophy has cut itself off the classical philosophical tradition, not because it has found what is wrong with the tradition or where it erred, but because it misunderstood the tradition. On the positive side, the purpose is to put forth the *Classical Theory of Mind* (CTM), as an account of mind and meaning in the classical philosophical tradition, and as an antidote to Analytic Philosophy. Chapters 1–6 will contain the polemical parts. Of these, Chapters 1 and 6 will contend against the Conservative Party and its leader Fodor; Chapters 4–5 will argue against the Radical Party, and Quine in particular; Chapter 2 and Section 5.5 will treat of Putnam's Middlebrow view; and Chapter 3 will look into the early steps — with Frege, Russell, and Carnap — Analytic Philosophy has taken away from term-based semantics and epistemology, and toward semantic and epistemic holism. The remaining chapters, 7–10, will contain the exposition of CTM, and proposals for solutions of some of its old problems. Chapter 7 will set out the formal or logical aspects of CTM, to be briefly sketched anon. Chapter 8 will deal with what we may regard as the material or nomological aspects; specifically, the issue of the implementation of CTM's posits in the brain, and of CTM as a natural science of mind. Chapter 9 will bring together and integrate the logical and the nomological parts into a unified account of mind and meaning, and develop the logical aspects further. Chapter 10 will put forward CTM's treatment of the so-called "Russell's paradox", using the logical means developed in Chapters 7 and 9. This is where the book will culminate; for the alleged paradox is one of the most characteristic features of Analytic Philosophy, whilst it has a clear and simple solution in CTM; the problem therefore neatly separates CTM from the Analytic tradition.

Here I will firstly outline the formal structure of CTM, and then raise the issue of the material implementation of CTM's posits. The mind comprises:

(α) A finite basis of semantically simple, *a posteriori*, empirical mental terms, or *ideas* (concepts).

(β) A generative mechanism, laden with finitely many semantically simple *a priori* ideas, for the production of *complex ideas* from the empirical basis.

(γ) A generative mechanism, laden with further finitely many semantically simple *a priori* ideas, for the production *sentences* from the simple and complex terms.

(δ) A finite basis of cognitive and emotive operations on tokens of the sentences, or *propositions*.

(ϵ) A generative mechanism for the production of complex cognitive and emotive operations on propositions.

A *mental state* is then construed as (an instance of) a cognitive or emotive operation on (a token of) a mental sentence, or proposition; a mental *process* is construed as a sequence of instances of such operations. The symbolic system (α)–(ϵ) is assumed to be organised in, and constitute, a single *psychic cell*; and the mind is assumed to be a complex system of such psychic cells. I will next point out the main semantic and epistemic features of CTM:

(ζ) The meaning of a public symbol derives from that of the mental symbol, an idea or proposition, which the public symbol is used to express on such and such an occasion.

(η) The meaning of a mental symbol consists in that the symbol represents or denotes a *nominal universal*; in the case of ideas, it consists in denoting a nominal *property*; in the case of propositions, it consists in representing a nominal *state of affairs*.

(θ) Meaning is *term-based*, rather than either sentence-based or holistic; the meanings of complex terms and sentences are built from the meanings of their constituent parts.

(ι) The mind immediately represents, or means, only its nominal world; whether, and to what extent, it veridically represents the real or noumenal world is an issue belonging to epistemology, not semantics.

(κ) There is a distinction between *a priori* and *a posteriori* knowledge, though not construed as in Analytic Philosophy. The distinction is a matter of which symbols — *a priori* or *a posteriori* ideas, or other propositions — are used as evidence in the evaluation of the proposition known. *A priori* knowledge is such that the evidence for the proposition known is drawn solely from the proposition's simple *a priori* constituent ideas, with *a posteriori*, empirical ideas being either absent or merely incidental to the evaluation of the proposition. (More generally, *a priori* knowledge is such that the evidence for the proposition known is drawn *solely from the mind's full range of simple* a priori *ideas*, not necessarily from the proposition's *constituent a priori* ideas alone; but the general case I will not deal with in this book. See a similar qualification in Chapter 7, Section 7.4.3.) *A posteriori* knowledge is such that the evidence is drawn *not solely* from the proposition's simple *a priori* constituent ideas, but also from other sources: either from its constituent simple *a posteriori* ideas (in addition to its *a priori* ideas) and nothing else, in which case we have *a posteriori observational* knowledge; or from sundry other propositions which may come from any branch of knowledge (in addition to its *a priori* and *a posteriori* constituent ideas), in which case we have *a posteriori holistic* knowledge. Unless the *psychological* distinction between *a priori* and *a posteriori* ideas is drawn, the *epistemic* distinction between *a priori* and *a posteriori* knowledge cannot make any sense.

(λ) There is a distinction between analytic and synthetic propositions; but again, not construed as in Analytic Philosophy. The distinction is a matter of the method and direction of evaluation of a proposition. Analysis is a method beginning from a proposition, under an assumption of truth-value, and proceeding by breaking the proposition down to its semantically simple constituent ideas, both *a priori* and *a posteriori*; an analytically true proposition is one which is provable by the method of analysis, in that if we assume the proposition as false, and break it down to its simple constituent ideas, we can show that the simple constituent ideas do not have, under the assumption of value, clear and distinct semantic identity, and hence that the proposition cannot be false (*i.e.*, must be true). In contrast, synthesis is a method beginning from a proposition's simple constituent *a priori* and *a posteriori* ideas, and proceeding by putting the ideas together, or combining them into the proposition, under the guidance of certain evidence. If the evidence guiding the synthesis is drawn solely from the *a priori* ideas, with *a posteriori* ideas being merely incidental or wholly absent, then we have *a priori* synthesis; if the evidence is drawn not solely from the *a priori* ideas, but also from the *a posteriori* ideas (and nothing else), then we have *a posteriori observational* synthesis; lastly, if the evidence is drawn not solely from the *a priori* and *a posteriori* constituent ideas, relying on sundry other propositions apart from the proposition being synthesised, then we have *a posteriori holistic* synthesis. Notice that analysis and synthesis are *not mutually exclusive*; in fact, any logically necessary proposition is provable either by *a priori* analysis or by *a priori* synthesis. Also, the distinction between analytic and synthetic propositions is *not semantical* but *epistemic*; it is a matter of the direction and method of evaluation, either by breaking a proposition down to its simple constituent *a priori* and *a posteriori* ideas, or by putting it together from its simple constituent ideas. Unless the mind is construed as a symbolic system allowing of such breaking down and putting together of propositions, and unless the psychological distinction is made between *a posteriori* (empirical) ideas and *a priori* ideas, with the *a priori* ideas organising the simple *a posteriori* ideas into complex ideas and propositions, the distinction between analytic and synthetic propositions cannot make any sense.

(μ) Logical modality, implication, and deductive validity are explained ultimately in terms of *a priori* analysis and synthesis; certainly not in terms of truth in possible worlds, nor in terms of recalcitrant public convention, nor in terms of deducibility in a sentence-based, axiomatic system or a system of natural deduction. The key notion in the explanation is that the simple constituent ideas of a proposition, including a conditional proposition or argument, put a constraint on what truth-value the proposition can have; some propositions their simple constituent ideas constrain to be true, and these are necessarily true; some they constrain to be false, and these are necessarily false; and some they allow to be either true or false.

The foregoing outline of CTM, in its psychological, semantical, and epistemical aspects, will be spelt out in Chapters 7–9. Here we may take notice of the main problem areas of CTM: there is the *formal problem*, concerning the structure and function of the symbolic system of the mind; there is the *material problem*, concerning the natural implementation of the symbolic system in the brain; and there is the *problem of integrating the formal and the material aspects* of CTM. It might seem, from an Academic point of view, that one need not worry about these problems all at once; and this would be true of a well-established science of mind. But although CTM harks back at least to Plato and Aristotle, and versions of it have appeared throughout the classical philosophical tradition — with Augustine, Anselm, Aquinas, Ockham, Descartes, Locke and Kant, among many others — there has never been a workable account of its natural implementation (not surprisingly, since natural science was until recently too immature for anyone to contemplate marrying the two); and this was perhaps the foremost reason why CTM declined and fell into oblivion in the 19th and the present centuries. The classical philosophical psychology seemed unsuitable for the project of making psychology into a natural science; the more people came to know about the natural world, the less it looked plausible that the human mind, according to the classical picture, had a place in it. Hence we have the origins of the Radical movement in Analytic Philosophy of mind and meaning. Ludwig Wittgenstein, one of the forebears of the Radical-*cum*-Middlebrow Parties, puts the material problem thus:

> How does the philosophical problem about mental processes and states and about behaviourism arise? — The first step is the one that altogether escapes notice. We talk of processes and states and leave their nature undecided. Sometime perhaps we shall know more about them — we think. But that is just what commits us to a particular way of looking at the matter. For we have a definite concept of what it means to learn to know a process better. (The decisive movement in the conjuring trick has been made, and it was the very one that we thought quite innocent.) — And now the analogy which was to make us understand our thoughts falls to pieces. So we have to deny the yet uncomprehended process in the yet

> unexplored medium. And now it looks as if we had denied mental processes. And naturally we don't want to deny them. (1953: § 308)

Here as elsewhere, Wittgenstein has made quite a few conjuring tricks himself. When he says that 'the analogy which was to make us understand our thoughts falls to pieces', he has in mind two things. Firstly, he has in mind that the semantics for the natural language used to speak about mental states and processes is not what the traditional account of language tells us it is; however, in hinting at the traditional account of language, he refers not to the position of CTM (as outlined in clauses (α)–(μ)), but rather the Conservative position (as outlined in clauses *(i)–(ix)*), especially the Conservative principle of extension-meaning supervenience; CTM he never understood, and never discussed, yet he wants us to believe that classical mentalism falls to pieces because the Conservative position, with its insistence on extension-meaning supervenience, falls to pieces. Secondly, he has in mind that the nature of mental states and processes has been left undecided, and their material implementation uncomprehended and unexplored, which is correct; but he manœuvres us into believing that the material problem has no solution, and that psychology must give up on classical mentalism altogether, seeking instead a radical behavioural-*cum*-linguistic solution (*cf.* §§ 304, 309).

A fly has found a way inside a bottle, and now buzzing about between the glass walls, is unable to find a way out. Wittgenstein recommends that we pretend the classical private mind is an illusion (§ 311), and take the linguistic turn; this is rather like advising the fly to take a buzzing turn inside the bottle. The fly could get out if only, instead of buzzing about, it walked up through the bottleneck; not a whirr, not a buzz, just walk; once it is through, it can fly again. I will propose a solution to the material problem of CTM, which seems to be no less hard to accept, for contemporary philosophers and scientists of mind, than it would be for the fly to accept that its only way out to freedom is to walk through the bottleneck. Specifically, I will propose that the empirical basis of mental symbols, the generative operations for complex symbols and propositions together with the *a priori* symbols laden in them, and the psychological operations on propositions, are implemented in the brain at the genetic level, as patterns of expression of the genetic material within certain neural cells, which I have called "psychic cells". Neither synaptic connectionism, nor any emergentism coupled with synaptic connectionism, will do the job of implementing the mind; nor, for that matter, conventional computationalism of the Conservatives. The genetic level is the only material level that could implement the symbolic system of the mind; there is nothing else in the brain that could do it. As we shall see in Chapter 8, the genetic level is eminently suitable for the general cognitive functions of learning, memory, recollection, for rational and associative processes, and for the causation of behaviour; and, in respect of CTM, it is eminently suitable for the analytic and synthetic symbolic processes which underlie the mind's *a priori* and *a posteriori* knowledge.

Without the genetic hypothesis, CTM would have no prospect of maturing and bearing offspring as a natural science of mind; and without CTM, so I will contend, there would be no prospect for anything worth regarding as a science of mind. Neither the Conservatives, nor the Radicals, nor the Middlebrows will give us what we need. The mind will not turn out to be an extension-determining Turing machine, as the Conservatives believe; nor will it go away and be replaced with an extension-under-determining Boltzmann machine, as the Radicals believe; nor will it emerge from behaviours, behavioural dispositions and the environment, like Venus from the waves, as the Middlebrows dream. Classical philosophical psychology, from Plato to Kant, cannot be lightly dismissed, despite the modern and post-modern trends; the best way to proceed is to make sure we understand its results, and then look into the brain to see how the mind so construed could be naturally implemented. A solution to the problem of implementation will not be, of itself, a solution to the metaphysical problem of materialism *versus* dualism, or materialism *versus* non-physicalism of some sort. The problem is only that of the *natural implementation* of ideas, as tokens of mental symbols, and operations on ideas, in the brain, insofar as they clearly do have a natural implementation. An answer to this problem need not be an answer to all questions about ideas; in particular, it need not tell us what makes an idea such and such a form of consciousness, and in general what consciousness consists in; our project will be difficult enough without these questions.

Anyway, time to begin the story.

Acknowledgements

I thank the following publishers for their kind permission to quote:

From TWO DOGMAS OF EMPIRICISM by W.V.O. Quine. Copyright © 1953 by the President and Fellows of Harvard College. Reprinted by permission of Harvard University Press.

From ONTOLOGICAL RELATIVITY AND OTHER ESSAYS by W.V.O. Quine. Copyright © 1977 by Columbia University Press. Reprinted with permission of the publisher.

From TESTABILITY AND MEANING by Rudolf Carnap, *Philosophy of Science*, 3:4 (1936) & 4:1 (1937). Reprinted with permission of the University of Chicago Press.

From REASON, TRUTH AND HISTORY by Hilary Putnam. Copyright © 1981 by Cambridge University Press. Reprinted with permission of the publisher.

From MIND, LANGUAGE, AND REALITY by Hilary Putnam. Copyright © 1975 by Cambridge University Press. Reprinted with permission of the publisher.

From REPRESENTATION AND REALITY by Hilary Putnam. Copyright © 1988 Massachusetts Institute of Technology. Reprinted with permission of the publisher.

From WORD AND OBJECT by W.V.O. Quine. Copyright © 1960 by The Massachusetts Institute of Technology. Reprinted with permission of the publisher.

From PSYCHOSEMANTICS: THE PROBLEM OF MEANING IN THE PHILOSOPHY OF MIND by J.A. Fodor. Copyright © 1987 Massachusetts Institute of Technology. Reprinted with permission of the publisher.

From A THEORY OF CONTENT AND OTHER ESSAYS by J.A. Fodor. Copyright © 1990 Massachusetts Institute of Technology. Reprinted with permission of the publisher.

From INDETERMINACY OF TRANSLATION AGAIN by W.V. Quine. © Copyright 1987 by The Journal of Philosophy, Inc. Reprinted with permission of the publisher.

From PHILOSOPHICAL PROGRESS IN LANGUAGE THEORY by W.V. Quine, *Metaphilosophy*, 1:1 (1970). © Basil Blackwell Publishers Limited. Reprinted with permission.

Chapter 1

Conservative Rationalism I

1.1 Common-sense Psychology

Jerry Fodor's Conservative account of mind and meaning is in some respects similar to the Classical Theory of Mind. I will show that it differs from CTM in certain features which are characteristic of Analytic Philosophy in general, and that precisely because of these features the account fails. Much of this chapter will be given to expository groundwork and a variety of minor arguments; Chapter 6 will contain the crucial case against Fodor.

The fundamental claim of Fodor's mentalism is that scientific psychology should rest on common-sense psychological explanation. At a minimum, this says that:

(i) There are psychological states that may be identified as, for instance, believing, desiring, perceiving, remembering, intending, expecting, fearing, supposing, hoping, *etc.*, that so-and-so is the case; *viz.*, as propositional attitudes.

(ii) There are psychological laws subsuming the attitudes and the behaviours caused by them.

Among the essential features of the attitudes are that they have *semantic properties* and *causal powers*. In other words, an attitude may be said to mean, or express the semantic property, that so-and-so is the case; and the attitudes of an organism are sufficient to cause the organism to behave in certain ways or have some other attitudes, according to the psychological laws. For example, desiring that *P* and believing that *Q is required for P* is normally sufficient to cause one to act so as to bring it about that *Q*; utterances that *P* are typically caused by beliefs that *P* and intentions to communicate that *P*; perceiving that *P* normally causes one to believe that *P*; believing that *a* is *F* is normally sufficient to make one believe that there is something which is *F*; and so forth.

As regards semantic properties, Fodor holds the Conservative view that the meaning of an attitude depends on which particulars or aspects of the natural environment it represents, so that two attitudes cannot be synonymous unless they represent the same things or aspects of the environment; *i.e.*, unless they have the same truth-conditions. As regards causal powers, he holds that causality and meaning are attributed to one and the same thing, the attitudes and, derivatively, the behaviours (including linguistic behaviours) subsumed under the psychological laws; and that consequently the laws do a double duty: on the one hand, they govern the *causal interactions* among the attitudes; on the other hand, they govern the — so to speak — *logical interactions* among the attitudes' propositional contents. Often enough to make acts of thinking useful, the logical interactions are rational inferences, so that the corresponding causal chains of thought mirror the forms of rational argument and result in rational action or confirmation of belief. In a qualified sense, one might say that propositional attitudes have a *dual nature*: as mental *operations*, token attitudes are states of an organism that causally interact with other such states, environmental inputs, and behavioural outputs; as *propositional* mental states, the attitudes have a semantic identity, and their causal sequences have a logical form. This dual nature of mental states constitutes the main theme of Fodor's account; and his project in the philosophy of mind and meaning is best regarded as an attempt to answer the question: "[w]hat sort of mechanism could have states that are both semantically and causally connected, and such that the causal connections respect the semantic ones?" (Fodor 1987: 14). We shall next look at the empirical hypotheses Fodor proposes as a solution to this problem.

1.1.1 The language of thought.

A belief (desire, *etc.*) that *P* is a relation that occurs between an organism and a *token* of a mental representation that means that *P*. For example, "Jane believes that bats are birds" is true just in case Jane bears the belief-making relation to a representation, tokened in her mind, that means that bats are birds. The representations are not merely an unordered bunch of mental symbols, but rather constitute a language-like symbolic system, the *language of thought*, or Mentalese. The system has certain characteristic properties traditionally attributed to natural language. In particular, the language of thought is *generative*, in that it comprises:

(i) a finite basis of semantically *simple terms*;
(ii) infinitely many semantically *complex terms* generable by syntactic rules from the basis;
(iii) infinitely many *sentences* generable by further syntactic rules from the simple and complex terms.

One of the notorious features of Fodor's position is that the finite basis of semantically simple terms includes all *prima-facie lexical* terms; that is, terms corresponding, roughly, to the lexicon of a public language such as English. This is a very extreme view. However, Fodor has wavered throughout his career as to the size of the basis; in some variants of his account, he regards the basis as smaller but unspecified. We shall come to consider his reasons for these variations later in this chapter.

The infinitely many complex terms and sentences generable from a finite basis of simple terms make up the domain for *mental processes*.

Mental processes are causal chains of instances of *operations* — such as believing, desiring, *etc.* — on *tokens* of sentential representations. Clearly, not all representations are involved in mental causation at a time. The representations, taken as symbol types, are causally inert; they *become* causally efficacious only when tokened in an attitude-making operation or relation. It is the operations, not the representations operated on, that have causal powers in the first place; and it is the representations, not attitude-making operations, that have semantic properties in the first place.

The notion that the mind is a system of symbols and symbolic operations is not new; it is implicit, and sometimes explicitly stated, in most classical philosophy of mind and meaning; and it is mainly in this respect that Fodor's Conservative position agrees with CTM. Some of the historical precedents of this notion will be reviewed in Chapter 7.

1.1.2 The computer analogy.

A salient aspect of the psychological processes of an organism is that, *exceptis excipiendis*, the causal order among the mental operations involved in the processes so determines the logical order among their propositional objects that the organism arrives at true conclusions provided it begins with true premisses. The problem is to specify just what property of the language of thought and the attitude-making operations it is that makes *rational causation* nomologically possible. The requisite property, Fodor says, is *syntax*: "...our best available theory of mental processes — indeed, the *only* available theory of mental processes that isn't *known* to be false — needs the picture of the mind as a syntax-driven machine" (1987: 19–20). Here, in an outline, is the theory. We think of the mind and mental processes by comparison to computing machines. Thus, Mentalese is a computational (formal, syntactically specifiable) language *like* the machine language built into the hardware of a computer. The attitude-making relations are *like* the programmes (algorithms, functions) that operate on formulæ of the machine language. Mental sentences are *like* the formulæ of the language. The formal language together with the operations constitute the *computational system* of the mind. As formulæ of the system, mental sentences are *symbols*: they have both syntax and semantic content. As objects of the operations, tokens of mental sentences have causal properties relevant to the computation of determinate outputs from inputs.

The causal and the semantic properties of mental symbols converge upon their syntax. In regard to rational psychological processes, this convergence is supposed to work as follows. Think of the syntax of a symbol token as some or other physical property, such as shape of electromagnetic state, by virtue of which the token causally interacts with other such symbol tokens comprising the physically implemented computational system. Such a system can be studied in purely formal terms, abstracting from its physical implementation; and it is possible to interpret the system so that the semantic relations among its formulæ mirror the syntactic relations (as is standardly done with classical systems of formal logic). It is therefore possible that there be causal syntactic processes respecting relations of meaning among the symbols involved. This is just what is required for the project of vindicating common-sense psychological generalisations, and of explaining how organisms act rationally out of propositional attitudes; and it being the only account of rational causal processes not known to be false, it is assumed as the best working hypothesis.

In accord with the computer analogy, Fodor accepts the highly implausible view that the mind, as implemented in the brain or some other physical system, has a certain cognitive architecture: *viz.*, the *classical computational architecture* of the serial von Neumann machine, based on a central processor in which symbols are displayed (or tokened) and operated on, and in which rational causal processes occur serially (see (1987: 16–19, 139)). This implausible view makes for one of the main reasons why many contemporary philosophers of mind and cognitive scientists reject Fodor's position; and since they identify his position with classical mentalism, they reject CTM in general. Yet although this reason holds good against Fodor, it does not apply to CTM; we shall see an alternative architecture for CTM in Chapter 8.

It is now doubtless evident how the problem of determining the identity of psychological states and processes is to be approached in Fodor's theory. For any organism, or any physically implemented mental system, a belief (desire, *etc.*) that *P* is a computational relation occurring between the organism and a token of a sentential formula of the mental code that means that *P*. The explanations to be thus formulated define the nomological identity of mental states, not the meaning of "believes that", "desires that", and so forth. They do not specify the logically necessary and sufficient conditions for the semantic identity of expressions or representations referring to the attitudes. In general, the questions answered by such explanations concern the *nature*, not *concept*, of the mind. Notice also that the identity of the operations cannot be determined in the same way as that of the formulæ. The formulæ, as mental representations, are symbols. Each such symbol is sufficiently distinguished by its syntax or computational form (though not by its meaning; the identity of meaning is necessary but not sufficient for the identity of symbol). The operations, in contrast, are not symbols; they have no syntax and, except derivatively from the symbols, no semantic content; they are functions defined over the sentential formulæ of the language of thought. As such, the operations are distinguished by their functional or causal role in the computational mental system; in turn, the functional or causal role of an operation is determined by the psychological laws governing the system.

Fodor's next problem is — and this is where most disputations in Analytic Philosophy flare up — what aspect of the formula that *P* it is that makes it mean, or carry the semantic content, that *P*.

1.1.3 The referential theory of meaning.

I pointed out earlier that Fodor endorses the Conservative notion that the meaning of a symbol depends on which particulars or aspects of the natural environment it represents, so that symbols cannot be semantically identical unless they represent the same things or aspects of the natural environment (*i.e.*, unless they have the same *extensions* or *truth-conditions*); in other words, each symbol has its meaning constrained by the extension or truth-condition it represents; and symbols of the same meaning must represent the same extensions or truth-conditions. In fact, Fodor's position is that representing extensions or truth-conditions is all there is to meaning, at least as concerns *categorematic* terms (such as "red", "Jane", "cat", *etc.*) and sentences. Concerning *syncategorematic* terms ("some", "all", "not", "and", "is a", *etc.*), he advocates a restricted functional-role or causal-role semantics, according to which the meaning of a symbol is determined by its role *vis-à-vis* other symbols.

Further, every meaning, or semantic property, is expressed by some symbol or another, but many syntactically distinct symbols may express the same semantic property. Also, many distinct tokens of the same type of symbol express the same semantic property. Distinct minds express the same meaning whenever they token the same basic, semantically simple symbol, since basic symbols are type-identical and universal for the human mind; and they expresses the same meaning whenever they construct syntactically identical complex symbols, since syntactic identity of symbol-tokens is sufficient for their semantic identity. In short, syntactically distinct symbols *may* (though, of course, need not) be synonymous, but syntactically identical symbols *must* be synonymous. This modest principle makes or breaks any form of semantic mentalism, whether Fodor's version of it, or CTM. Chapter 6 will show that Fodor's version of mentalism cannot satisfy the principle of meaning-syntax supervenience; and Chapters 7–9 will show that CTM has no difficulty with it.

I will now set out Fodor's semantics for the mental code in greater detail, using the following conventions. Bold-face English expressions will stand for the corresponding mental representations; capitalised expressions will stand for their semantic properties, or meanings; expressions in braces (*i.e.*, {, }) will stand for extensions of categorematic terms; standard logical symbols $\exists$, $\forall$, $\neg$, $\wedge$, $\in$, *etc.*, will be used for syncategorematic terms; truth-conditions will be written as compound expressions of extensions and logical symbols. For example, the term **marigold** expresses the semantic property MARIGOLD, and refers to the extension {marigold}. Again, **John loves Jane** expresses the semantic property JOHN LOVES JANE, and represents the truth-condition $(\text{John, Jane}) \in \{v \text{ loves } w\}$. Accordingly, **John**

loves Jane and **Jane is loved by John** will be taken as different symbols having the same semantic content. I will also need to speak of non-semantic properties, as any universals other than the meanings of symbols, and these will be indicated by wedges (*i.e.*, <, >); for example, the term **marigold** will be said to refer to the property <marigold> (as well as to the extension {marigold}, and to the particulars belonging to {marigold}). The term "reference" will be reserved for the relation between a *term* and its *extension*, or the *property* defining the extension, or the *particulars* belonging to the extension; the term "representation" will be used generically either for the relation of reference, or for the relation between a *sentence* and its *truth-condition*.

I will divide Fodor's semantics into four parts, dealing with: *(i)* categorematic terms; *(ii)* syncategorematic terms; *(iii)* sentences; *(iv)* sentential modalities and modal terms. It will be clear that the semantics is not peculiar to Fodor; it is common, implicitly or explicitly, to standard formal logic of Analytic Philosophy, Fodor's contribution being chiefly to assume it for the mental code. However, sketching it in some detail will help us to understand just what Fodor's position is, how it depends on the broader scheme of Analytic Philosophy, how it differs from CTM (*cf.* Chapters 7 and 9), and how we might go about assessing it.

(i) Categorematic terms.

Singular terms, such names and definite descriptions, refer to *single particulars*, and have their meaning because of that reference; so **Pretzel** has the meaning PRETZEL because it refers to Pretzel. General terms or predicates refer to *extensions*; *e.g.*, **cat** has the meaning CAT because it refers to {cat}; **loves** has the meaning LOVES because it refers to the extension {*v* loves *w*}; *etc.* Fodor wants to say that the identity of an extension is fixed by a property all members of the extension have in common: {cat} is fixed by <cat>, {*v* loves *w*} is fixed by <*v* loves *w*>, and so on; and he allows that **cat** refers either to {cat} or <cat>, *etc.* One of his reasons is that terms such as **unicorn** and **goblin** would otherwise be synonymous, both referring to the empty set. But he can have it that **unicorn** refers to <unicorn>, and **goblin** to <goblin>; and he can try to account for the relations of reference in terms of *nomic* relations between **unicorn** and <unicorn>, and **goblin** and <goblin>, not *causal-historical* relations between **unicorn** and {unicorn}, and **goblin** and {goblin}, which cannot be had since the extensions are empty. We shall come to the *nomic theory of reference* later in this chapter (Section 1.3); for now, it is worth bearing in mind that the referential theory of meaning goes hand in hand with a certain account of reference, and — according to Fodor — this account must be nomic and counterfactual-supporting, not merely causal-historical.

(ii) Syncategorematic terms.

Terms like $\exists$, $\forall$, $\neg$, $\wedge$, $\supset$, $\in$, *etc.*, have their meanings defined not by their reference but their *functional role* — specifically, their truth-functional role — *in the context of a sentence*; some of these may be semantically simple and others constructed from the simples, but in all cases their meanings are their roles in sentential contexts. For example, the meaning of $\wedge$ is defined in the context of $(\alpha \wedge \beta)$ by saying that $(\alpha \wedge \alpha)$ is true *iff* both α is true and β is true; the meaning of $\exists$ is defined by saying that $(\exists\delta)\Gamma\delta$ is true *iff* there is at least one particular in the domain of interpretation of the mental code which is Γ; and so forth. Fodor points out that although such functional role semantics is contextual rather than referential, it does not slide to semantic holism since both the functional roles and the contexts are well defined; holism would follow only if the roles were indeterminate and the contexts arbitrarily large, perhaps as large as the entire symbolic system (see (1990b: 110–111)).

(iii) Sentences.

Mental sentences mean what they do since they represent determinate *truth-conditions*. Thus, **Pretzel is a cat** has the meaning PRETZEL IS A CAT since it represents the truth-condition Pretzel$\in\{$cat$\}$; **All bats are birds** expresses the meaning ALL BATS ARE BIRDS since it represents the truth-condition $(\forall x)(x\in\{\text{bat}\} \supset x\in\{\text{bird}\})$; **1+1=2** means what it does since it represents the truth-condition $(1, 1, 2)\in\{u+v=w\}$; *etc.* Finding out whether tokens of the sentences are true is a matter of checking whether what **Pretzel** refers to is in {cat}, whether each particular which **bat** refers to is in {bird}, whether (1, 1, 2) is in $\{u+v=w\}$, and so forth. In general, we may attribute it to Fodor, and to much of Analytic Philosophy of Conservative orientation, that the human mind is in a position to find out, *from meaning alone*, not only whether $1+1=2$, but also whether all bats are birds, whether Pretzel is a cat — and even, as we shall see later, whether water is H_2O, the Morning Star is the Evening Star, whether all and only creatures with a heart are creatures with kidneys, whether Jocasta is Œdipus' mother, and so on — *provided it knows the meanings of the constituent terms*. We shall have many occasions to return to this astonishing consequence of the referential-*cum*-truth-conditional theory, and to marvel at the new vistas it opens for Natural Philosophy and the public weal.

(iv) Sentential modalities and modal terms.

Somewhat reluctantly but nevertheless, Fodor accepts the *possible-worlds* construal of such sentential modalities as necessary truth, contingent truth, and of the corresponding modal terms. According to this construal, (a token of) a sentence is necessarily true when it is true in all possible worlds, contingent when it is true in some and false in other possible worlds, *etc.*; and a modal sentence, or modal sentential form such as $(\Box\alpha \supset \alpha)$, is necessarily true when it is true in all possible worlds of all arrangements of possible worlds: in effect, in all possible worlds of possible worlds. A valid inference from premisses $\alpha_1, \ldots, \alpha_n$ to conclusion β is then construed as such that β is true in all possible worlds wherein $\alpha_1, \ldots, \alpha_n$ are true; and a true implication **if α then β** is construed as such that β is true in all worlds wherein α is. For Fodor's purposes of constructing a naturalistic theory of mental states and processes, the important modality is *nomological* (or otherwise less-then-logical) necessity, possibility, *etc.*, especially nomological inference and implication; and he adopts the standard construal that β nomologically (or less-than-logically) follows from α whenever β is true in all nomologically or otherwise *nearby* possible worlds wherein α is true. I think Fodor is well justified in having doubts about the possible-worlds construal of modal properties, and about possible-worlds semantics for modal terms; but it seems he does not realise that insofar as modern logic and his computational mentalism stick to the referential-*cum*-truth-conditional theory of meaning, there are no alternatives: the possible-worlds theory is essentially a *quantificational* construal of the meanings of modal terms, quantifying over possible worlds like any other particulars; and in this it is inseparable from the referential-*cum*-truth-conditional theory, using the same resources but introducing a new sort of particular. Analytic Philosophy does offer alternative approaches to modality, but these would not be acceptable to Fodor. There is the view that a necessary sentence is one which is *deducible* from a set of axioms by rules of inference, the axioms being taken as necessarily true by fiat. But here the fundamental bearers of meaning and subjects of analysis are *sentences* rather than terms; the semantics and methods of reasoning are *sentence-based* rather than term-based, which would not accord with Fodor. There is also the view that necessary truth is just a matter of *epistemic centrality* and *immunity to revision*; but this account is part and parcel of semantic and epistemic holism, which to Fodor is, rightly, unacceptable. Finally, there is a term-based account of meaning and modality in CTM proper; I will set out this account in Chapters 7–9, and contrast it with the above three approaches characteristic of Analytic Philosophy. This account should appeal to Fodor, but it certainly rejects the referential-*cum*-truth-conditional theory of meaning, and with it much of Fodor's version of mentalism.

1.2 The Supervenience Chain

We now understand all key aspects of Fodor's account of mind and meaning, except for the nomic theory of reference, to be given in Section 1.3. Here we shall look at some of Fodor's fine tuning in response to objections. The account may be conveniently expressed as follows. On the one hand, the 'best available theory of mental processes' requires that propositional attitudes be seen as computational operations on syntactically specifiable Mentalese symbols. On the other hand, the referential theory of meaning says that the sameness of extension or truth-condition is a necessary (as well as sufficient) condition for the sameness of meaning; *i.e.*, that Mentalese terms and sentences determine their extensions and truth-conditions, and that is what their meaning consists in. The former is a requirement of the theory of rational causation, the latter is the Conservative notion of meaning. The two are jointly equivalent to what may be called "the supervenience chain" structuring Fodor's account. The chain links organisms with their environments *via* propositional attitudes thus: identical organisms must (nomologically must) instantiate the same computational system; and identical computational systems must carry the same mental contents; and identical mental contents must determine the same extensions or truth-conditions.

Problem: is the supervenience chain internally coherent? *I.e.*, is the requirement that extension supervene on content compatible with the requirement that the computational mind supervene on the brain (or some other physical system)? We shall review three cases to the conclusion that the chain is not coherent, and look at Fodor's answers.

1.2.1 Syntactically identical symbol-tokens with different extensions.

Putnam (1975): Suppose somewhere in outer space there is a Twin Earth which is just like Earth, except that in lieu of the chemical compound H_2O there is, on Twin Earth, the compound XYZ. Suppose also that XYZ is indistinguishable from H_2O under the normal circumstances that obtain on Earth and Twin Earth (for instance, under normal temperatures and pressures, so that the two could be distinguished only by laboratory tests in extreme conditions). Finally, suppose Earthlings have not yet discovered that water is H_2O; and, likewise, Twin Earthlings have not yet discovered that 'water' is XYZ (so ensuring that Earthlings do not differ from their Twin-Earth *doppelgängers* in any relevant respect). Since I and my *doppelgänger*, Twin-Me, have identical brain states, it follows, assuming the Fodorian supervenience chain, that we have identical psychological states; so we express identical semantic contents; and since we express identical contents, we represent identical extensions and truth-conditions (think about the same things). But we do not. Twin-Me thinks about XYZ, whereas I think about H_2O. It follows that supervenient mental states do not determine their extensions; and, honouring the principle that different extension implies different content, it follows that psychological states, whatever else they may be, cannot supervene on the brain (or any other physical system). Contrary to Fodor, propositional attitudes cannot be computational operations on the formulæ of a language of thought.

Burge (1979): Suppose Earth and Twin Earth differ even less, or in a different sense. They are physically identical, except that, on Twin Earth, the 'English' linguistic community uses the phonetic form "arthritis" to refer not only to inflammation of the joints (as English speakers do on Earth), but also to certain rheumatoid ailments of muscles and tendons (contrary to English usage). Suppose lastly that I, on Earth, mistakenly believe that "arthritis" signifies both the inflammation of the joints and the certain rheumatoid ailments of muscles and tendons. Since I and Twin-Me have identical brain states, it follows, assuming the supervenience chain, that we have identical mental states; so we express identical semantic contents; and since we express identical contents (believe the same things), we represent the same truth-conditions (our beliefs are always true or false together). But we do not. When I and my *doppelgänger* think the formula **I am having arthritis in my thigh**, his thought may be true, whereas mine can not. Hence we do not express the same contents, and do not have the same mental states. It follows that mental states, whatever else they may be, cannot supervene on the brain, and the mind cannot be a computational system of operations on Mentalese symbols.

1.2.1.1 Narrow and broad meaning.

Fodor's response to the Twin-Earth arguments has vacillated over the years. He even used to consider distinguishing between *narrow* and *broad* meaning, with narrow meaning identified by the functional-causal role of a symbol, and broad meaning by its reference. Later he rightly abandoned this approach. Functional-causal role semantics, when applied generally to all rather than only well-definable symbols (such as $\wedge$, $\neg$), leads to semantic holism; and holism leads to nihilism as regards *term-based* individuation of meaning, and as regards the Conservative notion of meaning as extension-determiner which Fodor holds onto. His (1987) response still preserves the narrow-broad distinction, but both narrow and broad meanings are construed, as he would say, *atomistically*. Narrow meaning is a kind of *semantic potentiality*, a matter of having syntactically specified symbols organised in a computational mental system, with a capacity to acquire the only real meaning there is, which is broad meaning. Broad meaning is referential and *externalist*; that is, fixed by external symbol-to-world nomic relations; further, it is referential *relative to context*, where context is a *relevantly local domain of interpretation*. As for the Twin cases, Fodor says that Putnam and Burge are 'bloody-minded' in supposing that our Mentalese symbols, such as **water**, should be interpreted with respect to a domain including both Earth and Twin Earth. The relevantly local domain of interpretation of our Mentalese symbol **water** is Earth; that of Twin Mentalese is Twin Earth; so our **water** refers to $<H_2O>$ (or $\{H_2O\}$), while theirs refers to $<XYZ>$ (or $\{XYZ\}$). Fodor notes that common sense does not worry about the context or domain of interpretation of mental states; it is ready, like Putnam and Burge, to think it infinite. But cognitive science, as sophisticated common sense, makes us appreciate that we are finite and think locally; if we dare liken ourselves unto the gods, our mental contents shall be wrecked by such sophisms as the Twin-Earth examples.

Fodor's (1987) response is reasonable, for it is plausible that any symbolic system be interpreted with respect to a domain (*cf.* formal semantics for quantificational modal logic); and that the Mentalese symbolic system be interpreted with respect to a finite, relevantly local domain. But Fodor (1994) is not happy with it. This is because there is a conflict between, on the one hand, the *theory of broad meaning as domain-relative reference*, and, on the other hand, the *nomic theory of reference*. According to the former, domain-relative reference makes meaning. According to the latter, as we shall see in Section 1.3, reference is in turn made by nomic, dispositional mind-world relations; roughly, **water** refers to <water> since all and only instances of <water> *would* cause tokenings of **water** under psychophysically optimal circumstances. The modality there is nomological; *i.e.*, given the possible-worlds construal, **water** refers to <water> since all and only instances of <water> cause tokenings of **water** under psychophysically optimal circumstances in all nearby nomologically possible worlds. The nomologically possible world containing Twin Earth does appear to be nearby, there being supposed few differences between Earth and Twin Earth. It follows, given the nomic theory, that <water> must include both <H_2O> and <XYZ>; *i.e.*, it must be <$H_2O \vee XYZ$>; and that **water** in our Mentalese, as well as in Twin Mentalese, refers to <$H_2O \vee XYZ$>. So Mentalese and Twin Mentalese turn out synonymous after all, despite the domain-relativity of meaning, as long as Fodor holds onto the nomic theory of reference. In short, Fodor's (1987) notion of broad meaning as domain-relative reference does not solve the Twin-Earth problem; and worse, if one allows that the Twin-Earth scenario is nomologically possible, there will be no telling what **water** in Mentalese refers to and means, given the nomic theory reference; for the Twin-Earth scenario can be varied *ad libitum*, bounded only by nomological possibility. This would render Fodor's Mentalese meaningless, and reduce his account of mind to, so to speak, conceptual rubble. There comes a time — it came for Fodor in (1994) — when a man has to pound the table and declare: 'This is how it is going to be! There is *no* nomologically possible world in which I have a Twin who is just like me but means <XYZ> when I mean <H_2O>! And as for XYZ, "... there isn't any, and there couldn't be any, and so we don't have to worry about it" (1994: 29)!' Such is the power of possible-worlds semantics in the hands of an Analytic Philosopher. (Yet, in the heat of the moment, Fodor annihilated one possible world too many; for it would have been sufficient for his purposes to make sure that the accessibility relation from the Twin-Earth world to ours be *asymmetric*. Is it too late? Perhaps we could travel into the past ... ?)

From the perspective of my argument against Fodor, it does not matter at all whether he assumes for his account of the mind a domain-relative referential meaning, as in (1987), or a domain-free referential meaning, as in (1994). It matters only that he assumes the referential theory of meaning and the nomic theory of reference, as I will show in Chapter 6. In the rest of this section, we shall look at some of Fodor's troubles proceeding from the apparent semantic complexity of many lexical mental symbols.

1.2.2 Complex symbols under-determining extensions.

In the Twin-Earth cases, Fodor's supervenience chain is under strain since the computational mind is kept syntactically constant whilst the environment varies. But the chain threatens to break up also because most lexical symbols appear to be semantically complex; and complex symbols, so it seems, would determine their extensions only if they comprised necessary and sufficient conditions for the membership in the extensions; which they rarely do. This trouble does not depend on varying the mind's environment, and would apply even if the Twin-Earth scenario were nomologically impossible. Nor does it depend on the nomic theory reference, or the referential theory of meaning; any psychology accepting the semantical view that the meaning of a symbol determines its extension or truth-condition will be liable to it.

Putnam (1975; 1988) provides a number of examples indicating that most lexical mental representations do not determine their extensions. Suppose John's representation of elms is the same as his representation of beeches; *viz.*, **common deciduous tree**. The representations are identical, yet the extensions of the kinds <elm> and <beech> still are, respectively, the set of all elms and the set of all beeches. Likewise, suppose Jane's representation of gold is **yellow precious metal**; this is clearly insufficient to determine the extension of <gold>. In general, it is evident that we cannot use our mental symbols, within or without a domain of interpretation, as effective extension-determiners.

Such examples presuppose that most mental representations are semantically complex descriptions that could vary from individual to individual (and for each individual from time to time), and that would determine their extension if and only if they comprised necessary and sufficient conditions defining the property which all and only members of the extension have in common; and since many representations — such as John's and Jane's representations of elms, beeches, gold, *etc.* — seem not to comprise such necessary and sufficient conditions, it appears to follow that sameness of representation does not suffice, in general, for sameness of extension; *i.e.*, that representations are not extension-determiners.

Fodor agrees that most mental descriptions do not comprise necessary and sufficient conditions determining their extensions; but he takes this to show not that representations are not extension-determiners, but only that one's theory of the determination of extension, or reference, cannot be *definitional*; it can and should be *nomic* and *dispositional*, or, at worst, *causal* and *historical*. This response is common to Fodor (1987), (1990a, b), and (1994); and it goes back as far as (1975). But in the earlier versions of his account, Fodor held that all *prima-facie* lexical mental symbols (*i.e.*, symbols corresponding, roughly, to the lexicon of a public language) must be semantically simple, unstructured, unlearned, and thus innate and universal for the human mind; a sort of *wordwide rationalism*. This he held because he took it that lexical symbols could not be meaningful unless they were extension-determiners; but they could not be extension-determiners by virtue of defining their extensions; so there had to be a non-definitional, causal or perhaps nomic account of reference for them; but such an account would apply only to semantically simple symbols, since complex symbols referred (if at all) by description, whilst simples by causation; so lexical symbols must be one and all semantically simple, however counter-intuitive this may seem. Fodor (1982: 110–113) and (1986: 8–9) tentatively considered the supposition that lexical symbols could be descriptions, but the foregoing argument in favour of wordwide rationalism always prevailed. In (1994), Fodor has it both ways: representations such as **water**, **elm**, *etc.*, are allowed to be complex descriptions, but their meaning is their reference, with reference accounted for by the nomic and dispositional theory, not by definition. It follows that pairs such as **water** and **H_2O**, **the Morning Star** and **the Evening Star**, **Jocasta** and **Œdipus' mother**, **creature with a heart** and **creature with kidneys**, *etc.*, are syntactically distinct but *semantically identical*; and that such sentences as **water is H_2O**, **the Morning Star is the Evening Star**, **Jocasta is Œdipus' mother**, **a creature with a heart is a creature with kidneys**, *etc.*, are *analytic*; so that if only had Œdipus thought hard enough on what he *meant* by "Jocasta", if only had he tapped into the right meaning, he would have known from meaning alone that Jocasta is his long-lost Mother, and the tragedy would have turned into quite a comedy; alas, cognitive science was barely conceived in those days! Whether this *old sorcerer's rationalism* is less extreme than the wordwide variety I leave to be considered. But the reasons why Fodor is taken to such ends are not difficult to understand: he wants it, in accord with his supervenience chain, that identical physical states imply identical mental syntax; and identical syntax implies identical meaning; and identical meaning implies identical extension; but the definitional theory of the determination of extension, or reference, is false; so a causal or nomic theory of reference must hold *for all lexical symbols*. At this stage in the sorcerer's apprenticeship, quite a bit of magic could already be contemplated. But he further considers whether the nomic theory should apply to semantically simple symbols only, or to complex symbols also. The former alternative leads to wordwide rationalism, which has indeed some thaumaturgic virtues, but falls short on the side of *causal powers*, as we shall witness anon. The latter alternative has all the charms needed: you see, if you accept that the nomic theory of reference applies not only to semantically simple symbols, but to complex symbols as well, you will be able to tell why beliefs about water and beliefs about H_2O may have different causal powers but identical meaning; why desires about Jocasta and desires about Mother may have different causal powers but identical meaning; *etc.*; which you will find a distinct advantage to your art.

1.2.3 Syntactically distinct complex symbols with identical extensions.

Fodor has to allow that such pairs of beliefs as that there is life on the Morning Star and that there is life on the Evening Star, that water is wet and that H_2O is wet, that Jocasta is a widow and that Œdipus' mother is a widow, *etc.*, although synonymous *according to the referential theory*, are nevertheless different beliefs since they obviously have different causal powers. The problem is, what distinguishes such synonymous beliefs, and accounts for their distinct causal powers. Fodor notes that syntactically distinct symbols can have identical meanings; *e.g.*, in public language, "bachelor" and "unmarried man" are syntactically distinct but may be said to be synonymous; and that, according to the computer analogy, the causal powers of a belief-state are determined by the syntactic form of its Mentalese sentence together with the functional role of the belief-relation. Hence he rules that beliefs, desires, *etc.*, are relations between a creature, a *mode of presentation*, and a *proposition*, where a mode of presentation is a token Mentalese sentence, and a proposition is the meaning of the sentence. Pairs of beliefs such as that there is life on the Morning Star and that there is life on the Evening Star, *etc.*, are synonymous since they express the same proposition, but differ in their causal powers since they are operations on different modes of presentation. Does this solve the problem of synonymous beliefs with distinct causal powers? Here are two reasons why it does not.

Firstly, two syntactically distinct complex symbols could be synonymous only if the differences between them were not a matter of being composed of different semantically simple symbols, but merely a matter of different organisation of the simple symbols in the complexes. For example, $(Fx \wedge Gx)$ and $(Gx \wedge Fx)$ are syntactically distinct but synonymous, since the differences between them are merely a matter of ordering the constituent conjuncts. But if two syntactically distinct complex symbols differ in their simple constituent symbols, then they must be semantically distinct. Even with pairs of sentences of the form α and $\neg\neg\alpha$, the sentences are strictly speaking semantically distinct, since the superfluous double-negation still carries a semantic content. *A fortiori*, when two symbols differ essentially in their simple symbols, such as $(Fx \wedge Gx)$ and $(Fx \wedge Hx)$, there can be no doubt they differ in their semantic contents. Considering such pairs of Mentalese symbols as **the Morning Star** and **the Evening Star**, **water** and **H_2O**, **Jocasta** and **Œdipus' mother**, *etc.*, there can be no doubt that these must differ not only in their syntax but also in their meaning, since they must comprise different semantically simple constituents. It follows that Fodor cannot resort, in order to account for differences in their causal powers, to different modes of presentation, or different syntax, of such pairs of beliefs as that there is life on the Morning Star and that there is life on the Evening Star, that water is wet and that H_2O is wet, that Jocasta is a widow and that Œdipus' mother is a widow, *etc.*, so long as he sticks to the referential theory of meaning according to which these pairs are synonymous; for the syntactic differences between them could not be merely accidental, but must be differences in their simple constituent symbols, thus rendering them semantically distinct. So Fodor's problem with causally distinct but synonymous beliefs remains unsolved; yet it would be easily solved provided Fodor abandoned the referential theory of meaning, and allowed that such pairs of beliefs are semantically distinct.

Secondly, we saw at the beginning of this chapter that Fodor's project of establishing a naturalistic philosophy of mind and meaning requires him to find 'what sort of mechanism could have states that are both semantically and causally connected, and such that the causal connections respect the semantic ones'; and his answer is that the requisite mechanism is a syntax-driven computing machine. But there is a conflict between accounting for the causal powers of mental states in terms of their syntax, and the referential theory of meaning; for many states which *widely* differ in their syntax and hence in their causal powers turn out, assuming the referential theory, synonymous; so causal connections will not respect semantic ones, contrary to Fodor's project. For example, there is a semantic connection between **Jocasta** and **Œdipus' mother** in that **Jocasta is Œdipus' mother** is analytic, according to the referential theory. However, as the story goes, when Œdipus is caused to marry Jocasta by his beliefs and desires, he does not *mean* to marry Mother; nor is his action *irrational*, as it would have to be if there were a semantic connection between **Jocasta** and **Œdipus' mother**. He acts quite rationally given the meanings of his beliefs and desires; but his meanings do not determine their extensions. In short, Fodor wants to account for differences in the causal powers of mental states by their syntax, which is fair enough; he also wants to ensure that causal-syntactic connections among mental states respect *referentially construed* semantic connections (at least whenever the semantic connections are very close), in order to explain how organisms can act rationally out of the contents of their beliefs and desires; but the two are incompatible. On Fodor's theory, the tragedy of Œdipus is simply the result of some computational-*cum*-logical errors on Œdipus' part; hardly a story Sophocles would have bothered to tell.

As regards *psychological laws*, Fodor surmises that these obtain because the world — and all nearby nomologically possible worlds — is so constituted as to sustain a Leibnitian *parallelism* or *harmony* between psychological processes and the external world; that is, the worlds are such that individual minds in them are sufficiently similar, both in their internal computational properties and in their external nomic relations to the worlds, as to function according to natural laws which are specifiable equally by the causal-computational powers of mental states and by the referential contents of the states. Fodor says that if it turned out the worlds do sustain such a harmony, this "would show beyond any serious doubt that Turing [*in re* computational properties] and Dretske [*in re* semantic properties] between them have solved the mind/body problem. The foundations of cognitive science would then be secure, and the philosophy of mind would have nothing left to worry about ..." (1994: 56). Blessed are the souls who shall enter Fodor's rationalist paradise! But it will not be you or me, I fear; for, examining the conditions of entry — that one abide, in thought and word, by the referential theory of meaning, and that one's computational processes be in harmony with nature in all nearby nomologically possible worlds — 'tis plain the chosen soul who would enter must be, with a bit of rationalist magic and in all nearby nomologically possible worlds, *omniscient*; abracadabra, the path of a sorcerer's apprentice is thorny!

My final and principal argument against Fodor will be more involved than the foregoing, aiming to show that Fodor's account of mind and meaning fails to satisfy what is perhaps the most general requirement for any version of semantic mentalism, namely, that all tokens of a syntactically type-identical mental symbol be semantically identical. An account which fails on this criterion cannot be mentalistic and term-based; if the syntactic entities comprising the mind's symbolic system are not capable of holding their semantic identity from one occasion of tokening to another — that is, from one individual mind to another or, within the same individual, from one time to another — then there is no sense in which the entities could be said to be bearers of semantic properties, and in which public symbols could be said to derive their meaning from mental symbols. The only remaining measure of the meaningfulness of a public symbol would then be its overt behavioural use; and public symbols would be the only symbols worth ontological commitment. Such an account would collapse to semantic behaviourism and hence, as we shall see in the Chapters 4–5, to behavioural holism. I will argue that this is indeed the end Fodor's Conservative rationalism meets; and it is perhaps not surprising; for computational mentalism supplemented by the referential theory of meaning and the nomic theory of reference really amounts to Skinnerian *dispositional behaviourism*, so to speak, *computerised*; both hold that we are stimulus-response machines, the one using external, the other internal symbols; so similar ends should be expected for both.

1.3 The Nomic Theory of Reference

To show that Fodor's account collapses to semantic behaviourism and hence to behavioural holism, we need to show, firstly, that the account cannot satisfy the principle that syntactically identical symbol tokens must be semantically identical, so reducing the internal mechanisms of an organism to mere dispositional mediators between external inputs and behavioural outputs, and shifting the onus of semantic individuation from the organism's internal states to its overt linguistic responses to stimuli; and, secondly, that semantic behaviourism, taken as a *reductive* theory of meaning, collapses to behavioural holism. I will proceed as follows. In this section, I will set out Fodor's *nomic theory of reference*, which is the ground on which the battle must be fought. The following section will formulate my thesis against Fodor in greater detail; and the final section will settle on a *nomenclature* for the rest of the book, concerning terms such as "proposition", "concept", and several other. Chapters 2–5 will discuss the empiricist side of Analytic Philosophy, showing, among other things, that reductive behaviourism leads to behavioural holism. Chapter 6 will resume the case against Fodor, and demonstrate the claim that, given his account, the sameness of computational form is not sufficient for the sameness of meaning, so that the account collapses to semantic behaviourism. Fodor's Conservative rationalism and Radical empiricist behaviourism will thus turn out to be two sides of the same Analytic Philosopher's coin, the one buying no less or more than the other; needless to say, with inflation rates high, the coin will have lost much of its value over the years.

Fodor (1987) proposes the following. For any type of organism, a symbol α of the organism *refers* to the property $<\gamma>$ (or, what comes to the same thing, to members of the extension $\{\gamma\}$), if it is nomologically necessary that all and only instances of $<\gamma>$ cause tokenings of α, given certain qualifications (to be specified). For example, the symbol **water** refers to the property $<H_2O>$ if it is nomologically necessary that, under the conditions to be specified, all and only instances of $<H_2O>$ (or members of $\{H_2O\}$) cause tokenings of **water**. (To token a mental symbol, according to Fodor's usage, is to deploy the symbol in a mental operation or process. Fodor often speaks of 'putting the symbol into an organism's belief-box', but any 'attitude-box' would do just as well.)

The qualifications are of two sorts. Firstly, the *all*-clause needs to be amended so as to specify the circumstances under which an instancing of a property is nomologically sufficient for a tokening of the corresponding symbol. Secondly, the *only*-clause needs to be amended so as to allow for misrepresentation and error. I will deal with the amendments in that order.

1.3.1 The four-phase nomic theory of reference.

There is a small class of *psychophysical* properties and corresponding symbols, such that an instancing of any such property is nomologically sufficient for a tokening of the corresponding symbol *under conditions specifiable by psychophysical laws*, according to Fodor. For example, position an intact organism, John, with respect to a red wall so that John faces the wall, his eyes are open, there is enough light, John is close enough to the wall, and so on; if so, John must, as a matter of psychophysical laws, token the symbol **red**, so Fodor claims. Psychophysics is the science that determines how close John has to be to the wall, what the illumination must be like, *etc*. In general, it determines the conditions for the class of psychophysical properties and symbols, such that under those conditions *all* instances of any such property must causally occasion, in point of psychophysical laws, a tokening of the corresponding mental symbol.

What is the class of psychophysical symbols and properties? Fodor does not suppose that the variance between psychophysical and non-psychophysical symbols and properties is epistemologically or ontologically principled. Rather, what symbols and properties count as psychophysical is a contingent matter, to be decided by psychophysics itself. Nevertheless, he does suggest that the best candidates are the so-called "observation" or "sensory" terms and properties, since these do seem to covary, under psychophysically optimal circumstances, in the manner required by the nomic account of reference.

Although the psychophysical basis of symbols need not consist of sensory terms only, Fodor does deny that it extends much further. In particular, he denies that such symbols as **horse** — more generally, representations of medium-sized objects — fall within the domain of psychophysical laws. The reason is that such symbols (even though in some sense innate, according to the wordwide version of his account) may not be available for the psychological processes of an organism. One may, for instance, betoken an image of a horse under psychophysically optimal circumstances, yet without representing the horse *as a horse*; *i.e.*, without applying the symbol **horse** to the image. Further, as I mentioned earlier, Fodor denies that *prima-facie* lexical symbols are definable from any basis of semantically simple symbols, the psychophysical basis *inter alia*. If so, that is, if much of the Mentalese lexicon is neither included in the psychophysical basis nor definable from it, then what are the conditions for the nomic theory of reference in general? Fodor puts forth what we may regard as a *four-phase* account of the nomic connection between an instancing of a property and a tokening of the corresponding symbol. There is a *physical* phase, followed by a *psychophysical* phase, followed by a *psychological* phase, followed — for some symbols — by a *sociological* phase.

1.3.1.1 The physical phase.

Properties such as <horse> and <proton>, using Fodor's own examples, are not within the range of psychophysical laws. Nevertheless, such properties are nomologically responsible for instances of psychophysical properties under certain circumstances. In the case of <horse>, the circumstances are simple: being a horse is sufficient for having, as Fodor puts it, 'that horsy look'. In the case of <proton>, the requisite circumstances are more complicated: what is needed in order to link an instancing of <proton> with that of a psychophysical property is something like a proton detector, an experimental environment in which instances of <proton> have observable consequences. Under such circumstances, the causal dependence of the observable consequences on instantiations of <proton> is warranted by physical laws.

1.3.1.2 The psychophysical phase.

This begins where the physical phase terminates; and it maps, under psychophysically optimal circumstances, the observable consequences of instances of such non-psychophysical properties as <proton> or <horse> into the corresponding tokenings of *psychophysical representations*. The important point here is that such symbols as **proton** or **horse**, like the properties <proton> and <horse>, are not within the purview of psychophysical laws. The psychophysical phase ends in tokenings of psychophysical representations that correspond to the observable properties which occasion the tokenings; not, in particular, in tokenings of **proton** or **horse**. Hence the need for the following.

1.3.1.3 The psychological phase.

The role of this phase is to map tokenings of psychophysical representations into tokenings of the more theoretical symbols, such as **proton** or **horse**. This is done by means of *internalised theories* which organisms deploy in order to draw the necessary inferences from observational to theoretical representations. The use of internalised theories — as meaning-bearing symbolic entities — does not *ipso facto* invalidate the nomic account of reference as circular, so Fodor claims. This is because, firstly, the nomic theory requires only *that* there be a nomologically necessary causal connection between instances of <proton> and tokenings of **proton** under certain circumstances; it imposes no constraint on *how* the connection is mediated, so long as **proton** refers to protons because that connection obtains. In particular, it does not matter whether the connection runs *via* internal mechanisms which are themselves symbolic and so bearers of meaning. Secondly, Fodor suggests that we think of the internalised theories as computers between an organism's sensorium and 'belief-box', that take tokenings of psychophysical representations as inputs, and that output tokenings of the appropriate theoretical symbols. Regarding the internalised theories as computers takes away the need to refer to the meanings of the theories, and in general to the semantic properties of the psychological phase of the nomic connection between instances of a property and tokenings of a symbol.

1.3.1.4 The sociological phase.

For minds who are not dendrologists, for example, the psychological phase alone is not enough to ensure that tokenings of such psychophysical representations as one gets of elms or beeches are mapped reliably into tokenings of the corresponding symbols **elm** or **beech**, even under the best of psychophysical circumstances. These lay minds, Fodor says, rely on *experts* to make their tokenings of **elm**, **beech**, *etc.*, nomologically dependent upon instantiations of <elm>, <beech>, *etc.* Laity use experts to align reliably tokenings of their symbols with instancings of the appropriate properties, in the same way as experts might use a laboratory environment to align their tokenings of, say, **proton** with instantiations of <proton>. (This phase is not covered in Fodor's (1987) or (1990a, b) account, but it is prominent in (1994).)

In short, the four-phase nomic theory of reference is that a symbol refers to a property (or extension, or members of the extension) because there is a nomic chain that links instances of the property with instances of a psychophysical property in point of a physical law, that links the latter with tokenings of a psychophysical representation in point of a psychophysical law, and that links tokenings of the psychophysical representation with tokenings of the symbol in point of a psychological law. Lay minds, who for want of education or natural aptitude fail to acquire the requisite internal theories which sustain the psychological phase of reference, are not abandoned by the system, but required in point of a sociological law to defer to experts and other superiors, subsisting as it were on a referential dole; (sometimes they have to work for it, and sometimes it is cut off when they have no fixed address; but by and large everyone lives in harmony with nature and each other in all nomologically possible worlds).

1.3.2 An error in an account of error?

I will now turn to the amendment concerning the *only*-clause. The problem is that so long as a symbol refers to a property if all and *only* instances of the property are nomologically sufficient for tokenings of the symbol under certain circumstances, it follows that symbols cannot *mis*represent. For if both instances of the property $<\gamma>$ and instances of the property $<\delta>$ are nomologically sufficient for tokenings of the symbol α under certain circumstances, it seems that α should refer to $<\gamma$ or $\delta>$ rather than just $<\gamma>$ or just $<\delta>$. This is what Fodor (1987; 1990a, b) calls "the disjunction problem". The solution he offers follows the intuition that misrepresentation should be ontologically dependent on veridical representation, but not *vice versa*. Thus, Fodor suggests that since instances of $<\delta>$ do not belong to the extension of α, though instances of $<\gamma>$ do, we have it that, on the one hand, $<\delta>$-caused tokenings of α are not possible unless there exists independently a semantic set-up between instances of $<\gamma>$ and tokenings of α; whereas, on the other hand, $<\gamma>$-caused tokenings of α are possible regardless of there being any semantic relation between instances of $<\delta>$ and tokenings of α. $<\delta>$-caused α-tokenings are therefore ontologically dependent upon $<\gamma>$-caused α-tokenings, though not *vice versa*. In general, misrepresentation depends on the existence of a semantic set-up for veridical representation, but veridical representation is not so dependent on misrepresentation. Fodor says that misrepresentation is *asymmetrically dependent* on veridical representation; and the amended nomic theory of reference is that α refers to $<\gamma>$ if it is nomologically necessary that all instances of $<\gamma>$ cause tokenings of α under certain circumstances, and all tokenings of α caused by instances of $<\delta>$, for any $<\delta>$, are asymmetrically dependent upon $<\gamma>$-caused α-tokenings.

However, the emendation will not do. The reason is that Fodor tacitly presupposes throughout what he sets out to prove: namely, that tokenings of α which are nomologically dependent both on instances of $<\gamma>$ and on instances of $<\delta>$ do nevertheless refer to the property $<\gamma>$ rather than the property $<\gamma$ or $\delta>$; equivalently, that instances of $<\delta>$ do not belong to the extension of α, although instances of $<\gamma>$ do; or, again, that the extension of α is $\{\gamma\}$ rather than $\{\gamma$ or $\delta\}$. Short of begging the question of semantic individuation, there is no asymmetry between the nomic relation among instances of $<\gamma>$ and tokenings of α, and the nomic relation among instances of $<\delta>$ and tokenings of α, assuming as we do that under certain circumstances both instances of $<\gamma>$ and instances of $<\delta>$ would reliably cause tokenings of α.

I take it the nomic theory of reference still is without an account of misrepresentation, which points to a flaw in Fodor's pre-established harmony; nay, wrong symbols are tokened even in polite conversations, and cats play with mice in the best of all nomologically possible worlds, as Master Pangloss himself has explained to us ever so often, never so wisely.

1.4 *Casus Belli*

My claim that the sameness of computational form is not sufficient for the sameness of meaning construed referentially, and that computational mentalism *à la* Fodor will therefore collapse to semantic behaviourism and hence to behavioural holism, is a special case of a more general semantical claim, applicable to any version of mentalism, that the sameness of mental symbol, syntactic and semantic, need not be sufficient for the sameness of extension or truth-condition; that is, that mental symbols need not determine their extensions and truth-conditions. It is in this latter form that I will usually present the claim on the polemical side of the book, often expressing it as the denial of the Conservative *principle of extension-meaning supervenience*. This Conservative semantical principle has been, we may safely say, universally attributed by Analytic Philosophers to all versions of classical, term-based semantic mentalism. In fact, the disputations in Analytic Philosophy about meaning have drawn the sides so that, on the one hand, there is the orthodox view that meaning *determines* extension and truth-condition, or even that it *is* extension or truth-condition (a view attributed to traditional term-by-term mentalism); and, on the other hand, there is the unorthodox view that meaning is a matter of overt linguistic-

cum-behavioural use and convention, and as such *under-determines* extension and truth-condition (a view of the intellectually *avant-garde*, who pave the way for our future). Some Analytic Philosophers, such as Quine, are wholly on the side of the *avant-garde*; some, such as Fodor, hold onto the orthodox view; yet others, such as Putnam, have it pragmatically both ways, hoping for an emergence of the orthodox from the unorthodox. But all agree as to how the sides are to be drawn: it's the principle of extension-meaning supervenience *versus* that of overt behavioural use, or else straddling the two. I will suggest that in casting the disputation in this manner, Analytic Philosophy has effectively cut off its link — much to its detriment — to the classical philosophical tradition of, among others, Descartes and Locke, who have never held either side of the dispute. There is no question where my own allegiance lies; accordingly, the book is organised around four broad theses, three negative and one positive. This chapter and Chapter 6 argue against the Conservatives; Chapters 4–5 argue against the *avant-garde*; Chapter 2 and Section 5.5 expose the Middlebrows; and Chapters 7–10 do my best for the Classical Theory of Mind.

1.5 Nomenclature

I have so far avoided using the terms "concept", "idea", and other terms of art. This is because neither Fodor nor any modern or classical writer I know of use these terms with strict regularity, which has been a cause of seemingly endless confusions. It will serve to advance our subject if we settle the use of these and several other terms by the following conventions.

Both "concept" and "idea" will be used to stand for *tokens* of mental terms, as distinct from terms as symbol *types*. Similarly, "proposition" will stand for *tokens* of mental sentences, with sentences being symbol types. So concepts or ideas are constituent parts of propositions, whereas terms are constituents of sentences. The terms "symbol" and "representation" will be used generically to cover any symbol type or token, simple or complex. The term "notion" will be reserved to stand for semantic properties, or meanings, whether of symbol types or tokens. So **John loves Jane** and **Jane is loved by John** are different propositions, as symbol tokens, expressing the same notion JOHN LOVES JANE.

Notions are semantic universals or meanings, and are written in caps; non-semantic universals, as I stipulated earlier, are written between wedges: *e.g.*, <gold> is the property of being gold. However, I will now restrict this notation to apply solely to *physical properties*, or *natural kinds*, the sort of universal Locke called "real essences"; thus <gold> is the real essence of being gold. I will distinguish real properties from *nominal properties*, or *nominal essences*, as universals the identity of which is determined by symbols — more generally, by the mind — rather than by the physical environment itself; and these will be written between square brackets: *e.g.*, [gold] is the nominal property of being gold, the identity of which is fixed by the complex concept **gold** rather than the real nature of gold. These distinctions will suffice for the purposes of Chapters 2–9; in Chapter 10, it will be necessary to distinguish natural kinds, as physically real properties, from the following three sorts of real universal: *metaphysical kinds*, *psychological kinds*, and *social kinds*; the notation for these kinds will be introduced when it is needed.

The term "state of affairs" will be used for *nominal* universals which may be said to be represented by *sentences* or *propositions*; thus the proposition **John loves Jane** may be said to represent the state of affairs [John loves Jane]. States of affairs, as nominal universals represented by (tokens of) sentential symbols, will be strictly distinguished not only from real universals, but also from *truth-conditions*; just as nominal properties represented by terms will be strictly distinguished not only from real properties, but also from *extensions*. The term "extension" will stand for the set of things which instantiate a common *real* property: *e.g.*, the extension {gold} is the set of particulars that are gold, or instantiate <gold>; {gold} is not to be conflated either with the nominal property [gold], or the real property <gold>. Members of {gold} will be said to *instantiate* <gold>, and more or less *partake of* [gold]. Similarly, the term "truth-condition" will stand for a structure of things, or of *sets of* things, which instantiate certain *real* properties, and which may more or less partake of a (nominal) state of affairs. The truth-condition $(\text{John, Jane}) \in \{v \text{ loves } w\}$, for example, consists in that the ordered pair (John, Jane) belongs to the set $\{v \text{ loves } w\}$; and the truth-condition may more or less partake of the nominal state of affairs [John loves Jane]. Again, the truth-condition $\text{Pretzel} \in \{\text{cat}\}$ consists in that Pretzel belongs to the set of cats; and the truth-condition may partake of the nominal state of affairs [Pretzel is a cat]. The term "fact" will be used to stands for either the *complex real universals* which are or were instantiated in the environment, and which may be represented by propositions; or for the *truth-conditions* which do or did obtain in the environment. Thus, the complex real universal <water solidifies at 0°C> is a fact; also, provided John does love Jane, the truth-condition $(\text{John, Jane}) \in \{v \text{ loves } w\}$ is a fact, albeit psychological rather than merely physical fact. (We shall see in Chapter 10 that, apart from physical facts, there are metaphysical, psychological, and social facts; but there is no need to complicate the picture at present.)

The term "denotation" will be used to signify the semantic relation of representation between a *symbol*, whether type or token, and a *nominal universal*, whether property or state of affairs. In contrast, the term "reference" will be reserved for the relation between a *symbol* — again, whether type or token — and a *real universal*, or the *extension* or *truth-*

condition determined by the universal, or the *particular members* of the extension which instantiate the universal. So we shall say that the concept or idea **gold** denotes the nominal property [gold], and refers to the real property <gold>, as well as to (members of) the extension {gold}. The nominal property [gold] I will regard as the *denotatum* of **gold**, whereas the particulars which instantiate <gold>, and belong to {gold}, I will regard as the *referents* of **gold**.

The term "representation" will stand generically for either the relation of denotation or for the relation of reference, and will be used with either (tokens of) terms or sentences. With sentences or propositions, I will prefer to use the generic term "representation" to speak about the semantic relation between them and their *denotata* (states of affairs), and the relation between them and their truth-conditions; and this solely to avoid too great a discord with common usage in speaking of denoting or referring between a *sentential* symbol and what it represents. (It is perhaps worth mentioning that I certainly reject the Fregean notion that the referent of a proposition or thought is its truth-value; this has no place in the nomenclature.) Thus, I will say that **John loves Jane** represents the nominal state of affairs [John loves Jane], and also that it represents the truth-condition (John, Jane)$\in\{v \text{ loves } w\}$.

Finally, each concept or idea, as well as proposition, as a token of a mental symbol, will be said to have a discrete *form of consciousness*, either simple or complex; and this will be indicated by square double-brackets: for instance, the simple idea **yellow** has the simple form of consciousness ⟦yellow⟧; the complex idea **gold** has the complex form of consciousness ⟦gold⟧; and the proposition **gold is yellow** has the complex form of consciousness ⟦gold is yellow⟧.

The reader need not memorise these conventions; I will make every effort to clarify them in context. Nor is it necessary to regard them as anything but conventions; changes may be made as need be. Nor should they be expected to be wholly unambiguous and exhaustive; still, I will follow them throughout this book as closely as I can.

Chapter 2

The Idea as World and Will

A Belief in a Contribution of Environment to Meaning
and a Division of Semantic Labour

2.1 The Ambiguous Meaning of "Meaning"

In the decades after Kant, some people were misled to believe that the natural world is but an idea made more or less at one's will; since then, we have come a full circle, with some Analytic Philosophers today believing that an idea is constituted by the natural environment, and that it takes an expert's effort of will to apprehend it. This inverted idealism was anticipated by Kant: "... if it be really an objectionable idealism to convert actual things (not appearances) into mere representations, by what name shall we call that which, conversely, changes mere representations into things?" (1977: 955). Kant suggests that we call it "dreaming idealism", and perhaps that would be most appropriate; but we already have working titles for this view: "the extensional theory of meaning", and more recently, "the doctrine of the contribution of environment to meaning", due to Putnam (1975; 1988).

Further, if one is impressed by the extensional theory of meaning, or by the view that the natural environment itself contributes to, or fixes the identity of meaning, then — noting that on most occasions one is not able to determine the extensions of one's thoughts and speech — one will be inclined to believe that meaning is not quite within one's own power, and that others may be better in the business of meaning in their areas of expertise; in other words, that there is a *division of semantic labour*. One will be also inclined to believe that, ultimately, the natural environment itself sets the standards for meaning, regardless of anyone's attempts at meaning, whether one's own or any experts'. These melancholic humours were raised to be Analytic Philosophy by Putnam's Middlebrow Party.

2.1.1 Scientists as semantic experts.

Putnam claims to have brought to light the phenomenon of the division of linguistic labour. It is a fundamental socio-linguistic fact, he says, that the ordinary division of labour in a society engenders a corresponding division of the labour of knowing the meanings of such public symbols as "gold", "water", "elm", *etc.*:

> ... everyone to whom gold is important for any reason has to *acquire* the word 'gold'; but he does not have to acquire the *method of recognizing* if something is or is not gold. He can rely on a special subclass of speakers. The features that are generally thought to be present in connection with a general name — necessary and sufficient conditions for membership in the extension, ways of recognizing if something is in the extension ('criteria'), etc. — are all present in the linguistic community *considered as a collective body*; but that collective body divides the 'labor' of knowing and employing these various parts of the 'meaning' of 'gold'. (1975: 227–228)

Putnam goes on to say that the division of linguistic labour increases as the corresponding division of non-linguistic labour becomes more prominent; and that, with the rise of science, even words that previously were not subject to the division of labour — "water", for example — begin to manifest it. The critical point concerning the putative division of linguistic labour is that the various necessary and sufficient conditions for the membership in the extension of, say, "water", the different ways of recognising water, become "part of the *social* meaning of the word while being unknown to almost all speakers who acquire the word" (1975: 228). The moral Putnam draws hence, as regards any Conservative mentalistic account of meaning such as Fodor's, is that:

> Whenever a term is subject to the division of linguistic labor, the 'average' speaker who acquires it does not acquire anything that fixes its extension. In particular, his individual psychological state *certainly* does not fix its extension; it is only the sociolinguistic state of the collective linguistic body to which the speaker belongs that fixes the extension. (1975: 229)

Not all speakers have the same share of linguistic power in the socio-linguistic state which fixes the extensions of such words as "gold", "water", *etc.*; scientists, as *extension-determining experts*, get the lion's share, with average speakers deferring to them. In this account, the Conservative principle of *extension-meaning supervenience is preserved* at the expense of the principle of the supervenience of meaning-bearing psychological states on individual brains; not just *linguistic states*, but also *psychological states* become, so to speak, states of a collective meaning-bearing body. However, as we shall see next, this is not the only semantical position one can fairly attribute to Putnam (1975).

2.1.2 A Middlebrow compromise between extensionalism and mentalism.

On the one hand, Putnam (1975) rejects the *extensional theory of meaning*; this is the view that the meaning of a term *is* its extension, with extension being "simply the set of things the term is true of" (1975: 216). One reason he finds conclusive against the extensional theory is the creature-with-a-kidney/creature-with-a-heart example of Quine (1951). On the other hand, he rejects the traditional mentalistic theory of meaning; *i.e.*, the view (according to Putnam) that meanings are entities in the mind, and that "knowing the meaning of a term is just a matter of being in a certain psychological state" (1975: 219). His arguments here are the Twin-Earth examples, the elm/beech case, and the like. Putnam's own position on meaning is, in a sense to be specified, a *compromise* between the extensional theory of meaning, and what he believes to be traditional mentalism. The position is that the meaning of a word is an ordered pair of components comprising, as one of the components, the *extension* of the word, and, as the other component, the *stereotype* associated with the word. (More precisely, Putnam's proposal is to identify the meaning of a word with a quadruple of components: the extension, the stereotype, the semantic markers related to the stereotype, and the syntactic markers related to the word; the latter two need not concern us at present.) The extension of the word is, to reiterate, the set of things the word is true of; in contrast, the stereotype is the mentalistic component of the word's meaning. It is the minimum knowledge concerning the extension and use of the word, that would enable an individual to deploy the word in discourse and participate in the socio-linguistic labour of fixing the word's extension. The stereotype is a *mental description* representing certain salient and socially obligatory features of the extension; it does not, by itself, determine the extension of the word. Putnam's position needs therefore an independent account of how the extensional component of the meaning vector of a word is determined, and also an independent account of how the stereotype of the meaning vector is determined. This is where his beliefs in the division of semantic labour and the contribution of environment to meaning come into play.

According to the belief in a division of semantic labour, a variety of experts determine the extensions of such words as "gold", "water", "beech", *etc.*, and laity depend on them for the determination of these extensions. But, as Putnam points out, words such as "water" did not manifest a division of semantic labour prior to the rise of science, and words such as "pencil", "chair", "bottle", *etc.*, do not manifest it at all. One can ask, then, what the account of the determination of extension is for words which are not subject to the division of semantic labour. Putnam holds that such extensions are fixed *indexically*, by *ostensive definition.* For example, the extension of "water" may be fixed by focussing on normal samples of water and implicitly stipulating that nothing falls into the extension of "water" unless it is of the same nature or *kind* as those samples, whilst everything that is of the same nature as the samples does fall into the extension of the word. This is to say that the extension of such words as "water" is, in part, determined by the nature of the environment itself, whether or not that nature is fully known to any individual speaker who may be said, in a sense, to know the meanings of the words. The same applies not only to natural-kind terms such as "water", but also to terms referring to artifacts, such as "pencil", "chair", and so on. Putnam calls this "the contribution of the real world", or "the contribution of the environment" (to the meaning of a word).

> ... the extension of a term is not fixed by a concept that the individual speaker has in his head, and this is true both because extension is, in general, determined *socially* — there is a division of linguistic labor as much as of 'real' labor — and because extension is, in part, determined *indexically*. The extension of our terms depends upon the actual nature of the particular things that serve as paradigms, and this actual nature is not, in general, fully known to the speaker. Traditional semantic theory leaves out only two contributions to the determination of extension — the contribution of society and the contribution of the real world! (Putnam 1975: 245)

As regards the stereotype associated with a word, Putnam's position is that every linguistic community settles, for every word of the language, upon a standard minimum knowledge — which might be viewed as a *linguistic obligation* — concerning both the syntax and semantics of the word, which is demanded of each member of the community who is to be said to have acquired the word. The stereotype is that part of the linguistic obligation which bears on the semantics of the word, and which is required of each speaker who is to be said to know the word's meaning.

To summarise Putnam's (1975) position, the meaning of a word is an ordered pair comprising the extension of the word and the stereotype associated with it. The extension is determined, on the one hand, by a semantic labour divided among a variety of experts; and, on the other hand, by the nature of the environment itself. The stereotype is likewise socially determined, to be the minimum knowledge allowing an individual to participate in the social division of semantic labour. However, as we noted earlier, Putnam's view is not unequivocal. Putnam (1975: 227–229) speaks explicitly of the '*social* meaning' of, in particular, "water", rather than a 'meaning vector' comprising the extension and the stereotype associated with the word; and he says that this social meaning is a 'socio-linguistic state of the collective linguistic body'; he also includes various methods of recognising water in the social meaning of the word; still more at odds with his main position, he allows that — *via* the social division of semantic labour — even "the most recherché fact about water may become part of the *social* meaning of the word..." (1975: 228).

These data show that Putnam's meaning of "meaning" is ambiguous. Nevertheless, this much we may veritably attribute to him:

(i) that meaning determines extension;

(ii) that no individual mind, except an expert or a number of experts within their area of expertise, is able to determine the extensions of most words, and hence that individuals are not bearers of meaning;

(iii) ultimately, the environment itself determines what the extensions of words are, so that even an entire socio-linguistic body, with an advanced scientific culture and excellent research funding, may fail to know entirely the meanings of many of its words.

You will agree, it is a distressing semantics; rather than this, some would quit the principle of extension-meaning supervenience. But not the Middlebrows; in the spirit of pragmatic realism, they will have it both ways: the principle is true, and it is not; that is the answer ...

2.1.3 Pragmatic realism.

Putnam (1988) abandons both the *mentalistic* aspect of his former position (there is nothing, stereotype or whatever, in the mind of an individual speaker that helps to fix the meaning of a word) and the *extensional* aspect (the extension of a word is not a part of its meaning, though sameness of extension is still required for synonymy). Further, Putnam *virtually* dissociates the problem of meaning from that of the determination of extension, or reference; I say "virtually", since the account of meaning he offers comes with a certain account of truth, and truth is an essential feature of reference: reference amounts to *being true of* (1988: 1, 32); also, he endorses the Conservative rule that a difference in extension guarantees a difference in meaning, thus again linking the theories of meaning and reference (1988: 32, 34). More on the split of meaning and reference soon. We shall now look at Putnam's account of reference:

> ... it is difficult — I suggest, in fact, impossible — to give a *reductive* theory of reference. But if what we ask is not a reduction of the notion of reference to other notions regarded as metaphysically more basic, or a theory of "how language hooks on to the world," but simply a working characterization of how it is that words like "robin" and "gold" and "elm" manage to refer, then it is not difficult to give one. The fact is that some people know a good deal about certain kinds of things. These "experts" as I have been calling them may pick out these classes by different criteria. That doesn't matter as long as the criteria in fact pick out the same class. If experts in one country determine whether something is gold by seeing whether it is soluble in aqua regia and experts in another country determine whether it is gold by seeing whether it passes some other test, provided the two tests agree (or agree apart from borderline cases), then communication can proceed quite well. There is no reason to think of any one test as "the meaning" of the word ...
>
> But, it will be objected, this only accounts for how experts can use the word. However, there is no problem about how nonexperts can use the word: in doubtful cases, they can always consult the local experts! There is a *linguistic division of labor*. Language is a form of cooperative activity, not an essentially individualistic activity ...
>
> In sum, reference is *socially fixed* and not determined by conditions or objects in individual brains/minds. Looking inside the brain

> for the reference of our words is, at least in cases of the kind we have been discussing, just looking in the wrong place. (1988: 25)

Notice that Putnam's use of the term "reference" is ambiguous. Firstly, he uses "reference" to designate the semantic relation that obtains between a symbol and its extension, as in the phrase 'it is difficult to give a reductive theory of reference'. Secondly, he uses "reference" to designate what had better be called "referent" — *i.e.*, what is referred to by a symbol — as in the phrase 'reference is socially fixed and not determined by conditions or objects in individual brains/minds'. Thirdly, he uses "reference" to signify that which, putatively, fixes the extension of a word — *i.e.*, mental description or symbol, sociolinguistic state, *etc.* — as in the sentence 'looking inside the brain for the reference of our words is just looking in the wrong place'. This makes the semantic language-game all the more challenging.

It is clear that, contrary to (1975), Putnam (1988) does not regard such necessary and sufficient conditions for the membership in the extension of, say, "gold" as that gold is the element with atomic number 79, and such ways of recognising gold as the test of solubility in *aqua regia*, and so forth, as parts of the meaning of the word. He also states explicitly that it is not analytic that gold is the element with atomic number 79; that ways of recognising gold may differ from one group of experts to another and from time to time, so that none can be identified with the meaning of "gold"; and that "[t]he chemist who knows that the atomic number of gold is 79 doesn't have a better knowledge of the *meaning* of the word "gold", he simply knows more *about* gold" (1988: 23).

In accordance with (1975), Putnam (1988) upholds his belief in the contribution of environment to meaning, the view that the extension of such words as "water", "gold", "cat", "milk", *etc.*, is in part determined by the nature of the environment itself, and that what aspect of the environment fixes the extension is in turn determined by the nature or kind of the samples we focus on — or 'point at' — when using these words, regardless of whether or not anyone knows what the nature of the samples is. In short, Putnam maintains the indexical account of the determination of extension, and the account in terms of a division of semantic labour: the extension of a word is determined in part by scientists as semantic experts, and in part by the nature or kind of indexically fixed aspects of the environment (1988: 30–34).

Both of these accounts of the reference of a term such as "gold", although mentioning more or less explicitly various psychological descriptions associated with the term, or such descriptions stated in words, are given without connoting the term's meaning. Putnam says: "... the effect of my account ... is to separate the question of how the reference of such terms is fixed from the question of their conceptual content" (1988: 38). ("Conceptual content" should be read in the sense of "meaning", in this context.) Yet despite this, *mirabile dictu*, Putnam still holds onto the Conservative principle that a difference in extension requires a difference in meaning. This creates a little quandary in his pragmatic realism, as we shall witness anon:

> If we can account for how our words refer to the things they do without appealing to the idea that they are associated with fixed "meanings" which determine their reference, then why should we have such a notion as meaning at all? But this is not really such a puzzle: the best way to get along with people who speak a different language — or, on occasion, even to get along with people who speak one's "own" language in a different way — is to find an "equivalence" between the languages such that one can expect that — after due allowance for differences in beliefs and desires — uttering an utterance in the other language in a given context normally evokes responses similar to the responses one would expect if one had been in one's own speech community and had uttered the "equivalent" utterance in one's own language. As a definition of sameness of meaning this would not satisfy a skeptical philosopher like Quine: it would not satisfy him because, for one thing, the identification of contexts as "the same" presupposes the very "translation scheme" which is being tested for adequacy, and because the identification of beliefs and desires likewise presupposes translation. But in the real world, our problem is not the theoretical problem of "underdetermination" — the problem of the existence of alternative schemes which satisfy the criterion of adequacy equally well — but the difficulty of finding even *one* which does the job. That we do succeed in finding such schemes in the case of all human languages is the basic anthropological fact upon which the whole notion of "sameness of meaning" rests. (1988: 25-26)

The key features of Putnam's (1988) semantics are all implicit in this passage: namely, the abandonment of the Conservative notion of meaning as extension-determiner; the adoption of a *holistic*-cum-*behavioural* (though not *behaviouristic* — he rejects every form of semantic reductionism) notion of meaning as that which is preserved in translation; and the adoption of the maxim of allowing for differences in beliefs and desires, or *charity in interpretation*. These topics will be dealt with in detail in Chapters 4-5. For our present purposes, we need to understand only that, according to this position, the meaning of a symbol is a matter of its overt use, in the context of a larger symbolic network, *vis-à-vis* various stimulatory conditions both verbal and non-verbal; and that such holistic meaning is not a discrete semantic *property* of the symbol, but rather something unclear and indistinct, an *emergent property* which is not sufficient to determine the extension of the symbol. Whatever it is, we may be certain that — given the Twin-Earth scenario of Putnam (1975), preserved in (1988) — it is *identical* for us on Earth and for our *doppelgängers* on Twin Earth; so that, granted this position, as I on Earth and my *doppelgänger* on Twin Earth simultaneously utter the phonetical form "water", our utterances have *identical meanings*.

Notwithstanding this, Putnam also says that, supposing some of us had visited Twin Earth (or some of *them* had visited Earth) in 1750, before we discovered that our water is H_2O and they discovered that their water is XYZ, and supposing the visitor had had a chat about water with the natives, "[n]o one on Earth or on Twin Earth would have noticed that the word had a different meaning ..., but in my view they *would* have had a different meaning" (1988: 31–32). A veritable contradiction? Come now, be a *realist*, look at it *pragmatically*!

2.2 The Will to Linguistic Power

There are, as we have seen, three versions of the belief in a division of linguistic labour, each with a notion of the contribution of the environment to meaning. The original version is rather *oligarchic*; it does say that meanings are 'sociolinguistic states of collective linguistic bodies', but most of the linguistic power is concentrated in the hands of scientists regarded as semantic experts, laity having little or no input into the semantic business of determining extensions. The next version, which says that meanings are ordered pairs of extensions and stereotypes, introduces as it were a *Westminster system* of sharing linguistic power: whilst common people mean by stereotypes, the *élite* mean by extensions and hold final semantic authority. The last version, underlaid by semantic holism, is markedly *egalitarian* and *democratic*: everyone gets an equal share of meaning power, so long as one's linguistic and behavioural proclivities do not overtly deviate from the official semantic line (in which case one is rendered meaningless). This last version, as became obvious at the end of the previous section, is but a step removed from the notion that the natural environment has an absolute power to determine meanings, whether or not any individual knows what the meanings are; a semantic *tyranny*, so to speak. We shall now look into the respective merits of these systems, with a view to finding whether there is any justice in them.

2.2.1 Semantic oligarchy: scientists as unrepresentative leaders. This version says that the meaning of, in particular, "gold" comprises the ways of recognising whether or not something is gold, or necessary and sufficient conditions for the membership in the extension {gold}; and since only experts know such necessary and sufficient conditions, since average speakers need not know any of the ways of recognising gold, it seems to follow that there must be a division of semantic labour: the meaning of "gold" must be a collective socio-linguistic state, with most or all semantic authority in the hands of the experts, and laity dependent on them. Underlying this position is a belief that meaning is a matter of determining extensions, and verifying whether something does or does not fall into the extension of a word; which is something only scientists can do for many words in common use, and for which laity must rely on the scientists.

Clearly, though, Putnam conflates semantics *as a study of symbols*, with natural science *as a study of the environment*. Natural science attempts to determine the *nomological* identity of the property <gold>, and hence the extension {gold}; but that is an altogether different project from determining the *logical* identity, or meaning, of the symbol "gold". In natural science, we take it we already know the meaning of "gold", and endeavour to characterise the nature of the kind referred to by "gold" — *viz.*, the property <gold>, or extension {gold} — not presuming thereby to improve on the meaning of the term. In semantics, we endeavour to characterise the symbol "gold" in respect of its meaning. Without doubt, there is a division of scientific or, in general, *epistemic* labour as concerns such properties as <gold>; but the claim that there is a division of *semantic* labour as concerns the symbol "gold" would follow only if the meaning of "gold" were to be identified with the various methods of recognising, or *verifying*, whether something is gold; that is, only if semantic verificationism were true.

There are many good arguments against semantic verificationism, some of which we shall discuss in Chapter 3. Here we should note, firstly, that in (1988), Putnam himself has abandoned verificationism, saying that 'the chemist who knows that the atomic number of gold is 79 doesn't have a better knowledge of the *meaning* of the word "gold", he simply knows more *about* gold' (*op. cit.*); secondly, that in (1975), whilst identifying the meaning of a word with a collective sociolinguistic state, he also allows that individual speakers are able to mean on their own. Thus, Jane's concept of gold is **yellow precious metal**; what she *has in mind* — what she *means* — when uttering "gold" is **yellow precious metal**. In other words, Jane's utterance expresses her *idea* of gold, which happens to be of a yellow precious metal. Similarly, what John *has in mind* when uttering "elm" or "beech" is the idea **common deciduous tree**. These lay ideas do not differ, *insofar as the study of symbols is concerned*, from a scientist's idea of gold, say, **the element with atomic number 79**. In either case, the idea is a semantically complex description which more or less falls short of specifying the *real essence* of gold — the real property <gold> — and succeeds at best in specifying a *nominal essence* of gold: *e.g.*, [yellow precious metal], [the element with atomic number 79], *etc.* The nominality of the properties represented by Jane, John, or the scientist is guaranteed simply by the fact that their ideas are *complex*, and as such determined by the mind itself, not by the environment.

The extent to which one's nominal property [gold] coincides with the real property <gold> depends on whether one's complex idea of gold is veridical. The classical theorists differed much in their views concerning the mind's ability to represent real essences. Locke held that the mind can do it insofar as its semantically simple ideas represent essences which are *both real and nominal*, and some of the simple ideas — those standing for primary qualities — are veridical resemblances. Kant held that all our ideas, simple or complex, represent only nominal properties, and therefore that real or *noumenal* properties are absolutely beyond the mind's cognitive reach.

As regards semantics, we need not worry about this issue. The crucial point is that, even according to Putnam (1975), the bearers of meaning had better be *individual speakers*, with the symbols tokened in their minds, not 'collective sociolinguistic bodies', or experts, or learned societies. Although *complex* concepts may differ greatly from one individual to another, and from time to time for each individual, this does not give any special powers to scientists as concerns *meaning*. Scientists have special knowledge, but their ability to mean is not superior to anyone else's; their resources of simple ideas are the same, and in forming complex ideas they are on the same footing as laity, in that with vast majority and perhaps all of their complex ideas they represent only *nominal* properties (which could differ from person to person and from time to time). To put it in my nomenclature, scientists, just like laity, *mean* in that they *denote* nominal properties; both may and do succeed, as a matter fact, in *referring* to pieces of, say, gold, and to the real essence of gold; but what they *mean* consists not in referring to particular pieces of gold, nor in referring to the extension {gold}, let alone in referring to the real essence <gold>; it consists in that each *individual* scientist *denotes* some nominal essence [gold], which may differ from person to person and from time to time. In short, meaning is not referring but denoting; as concerns referring or determining extensions, scientists have indeed greater powers than laity; but as to denoting, all minds are equal individuals, each denoting its own nominal essences, meaning — so to speak — its own nominal world.

One might ask then, what is *the* meaning of "gold"? Within a linguistic community, there is a uniformity of complex ideas bound by a common public word (see Chapter 9). But complex ideas are only roughly regimented even in the most orderly of linguistic communities, so that the meanings of words used to express them have individual nuances. None of this, however, justifies the claim that there is a division of semantic labour, and that meanings are distributed throughout a society, with scientists being vested with special semantic powers; that view is a consequence of the wrong principle of extension-meaning supervenience, which holds that meaning *is* or *requires* determining extensions, or referring. Meaning is denoting, and denoting is *individualistic*: it might be that experts are able to form more useful complex ideas, and thus denote more useful nominal properties, as regards knowing the nature of the environment; but that does not make them better in *meaning*, though it may in *knowing*.

2.2.2 A Westminster system? Extensions as Lords, stereotypes as Commons.

According to this version, the meaning of a term is an ordered pair comprising the *extension* of the term and the *stereotype* associated with it. The stereotype does not, of itself, determine the extension; it is a minimum knowledge concerning the extension, allowing an individual to use the term in discourse and participate in what Putnam takes to be the semantic labour of determining the identity of the term's extension. The alleged division of semantic labour thus functions to ascertain the identity of the extensional component; but what the extensional component is, objectively, is settled by the nature of the environment itself.

In this system, laity enjoy more linguistic power than in the oligarchic system: they know a half of a word's meaning, the stereotype; still, the full measure of meaning is definitely beyond their reach, unless they be raised to the class of those who can determine extensions by knowing the right necessary and sufficient conditions. We have, however, already noted that knowing the nomologically necessary and sufficient conditions for the membership in the extension of, say, "gold", is an altogether different problem from that of knowing the logical identity, or meaning, of the symbol "gold" as used or tokened by this or that individual on this or that occasion; and that, to answer the latter problem, one need not know the answer to the former. This is just to reiterate that semantics is not simply the sum-total of natural science, and that natural science as such does not solve problems in semantics; accordingly, the division of scientific or, in general, epistemic labour is not a division of semantic labour.

I suspect, it was popular grievances such as these, against verificationism, which eventually led Quine and others (even Putnam, in his Middlebrow way) to resort to a semantical revolution: *the Conservative principle of extension-meaning supervenience must be overthrown*! Yes, I agree; but when it comes to the slogan "Meaning is Use!", I am afraid the baby has been thrown out of the tub with the bath-water.

2.2.3 Колхоз semantics, or meaning holism.

This version of the belief in a division of semantic labour, due to Putnam (1988), holds that such necessary and sufficient conditions for the membership in the extension of, say, "gold" as that gold is the element with atomic number 79, such ways of recognising gold as the test of solubility in *aqua regia*, *etc.*, are not parts of the meaning of "gold", so contradicting the first version, according to which the ways of recognising gold and the various necessary and sufficient conditions for the membership in {gold} are included in the social meaning of the word; it also contradicts, more indirectly, the second version, according to which the labour of determining the extensional component of the meaning vector of a word may be said to be a part of the semantic labour of knowing the meaning of the word. Contrary to these versions, Putnam (1988) emphasises that the ways of recognising gold differ from one group of experts to another and from time to time, so that none can be identified with the meaning of "gold"; and that 'the chemist who knows that the atomic number of gold is 79 doesn't have a better knowledge of the *meaning* of the word "gold," he simply knows more *about* gold' (*op. cit.*). But then, what could the alleged division of semantic labour consist in? For there is indeed a division of *referential* labour, the labour of determining the extensions of words; and scientists are indeed better than laity in fixing the extensions of words within their areas of expertise; but Putnam has just conceded that meaning is not referring, and a division of referential labour is not a division of semantic labour.

This is further exasperated by the fact that the semantical theory of Putnam (1988), underlying the claim that there is a division of semantic labour, is meaning holism. The trouble is that since holism dissociates the meaning of a word from its determination of extension, or reference — since, according to holism, meaning *under-determines* extension — it follows that the experts who determine the extensions of such words as "gold", "water", "elm", *etc.*, do not have a better knowledge of the meanings of the words; expert knowledge is not a part of the words' meaning. In other words, holism is *incompatible* with the belief that scientists are experts on meaning, and that there is a division of semantic labour.

Holism rejects the principle of extension-meaning supervenience; it does that mainly in order to ensure that laity, whose meanings in general are not extension-determiners, receive — as it were — a full share of linguistic power. But holism achieves semantic equality at a cost which is hardly worth paying: here, meaning becomes overt behavioural convention within a large socio-linguistic context, a kind of abject social conformism; semantic power is given back to the people, yet *individual* people really do not mean anything at all; they only *behave* as though they do, responding in much the same way to much the same stimuli.

2.2.4 Semantic tyranny: Merlin knows and rules by meaning alone.

Seeing that all stimulus-response machines are of roughly equal stature as regards meaning, it is not surprising that Merlin, who pretends to have a thaumaturgic access to *de re* significance, soon finds himself represented by the *media* (versified, sculpted, incantated, information-super-highwayed) as taller than the rest by the head. For, whilst common stimulus-response machines are semantic equals, in that they *would* respond roughly equally to roughly equal stimuli, Merlin is more equal than others: *the environment itself contributes to his meaning* (some say, in all nomologically possible worlds)! When Merlin utters your name, he does not merely *use* the symbol *vis-à-vis* sundry stimulatory conditions, in a feeble attempt to mean, he means *you*; and if your name be "Cicero", you cannot escape Merlin's justice by calling yourself "Tully" (especially if you are a patrician); Merlin will find out that "Cicero is Tully" is analytic, and you, the referent of your two names, shall be fixed solely by his meaning; which could not be otherwise, considering that you — a mere parcel of the natural environment — yourself contribute to Merlin's meaning, like it or not! Here we have the final stage in our progress toward semantic justice.

But does Merlin really have a thaumaturgic access to *de re* significance? We can examine his secret by critically dissecting his magic formula. He says that the extensions of such words as "water", "Cicero", *etc.*, are fixed *indexically*. That is to say, the extension of, *e.g.*, "water" is fixed by focussing on normal samples of water and implicitly stipulating that nothing falls into the extension of "water" unless it is of the same nature or kind as those samples, whilst everything that is of the same nature as the samples does fall into the extension of the word. In this manner, the extension of "water" is determined by the nature of the environment itself, whether or not that nature is fully understood by any individual speaker who may be said, in a sense, to know the meaning of the word. Thus the word's meaning depends on the nature of the environment: the environment itself contributes to the semantic identity of the word. Putting on a wizard's cap, it becomes clear that anyone who knows the formula, and taps into the right *de re* meaning, should be able to accomplish quite a bit of magic: find out how the natural world is solely by semantic analysis, and who knows even make it as one pleases by meaning as one pleases. Much better than an atom bomb! With this, Merlin rules the world!

On second thoughts, though, behind the magic there seems to be a rather simple trick. We can distinguish two spell-binding phases in Merlin's thaumaturgic formula: firstly, there is the phase of focussing on (indexing, pointing at) certain samples of the substance represented by the word; secondly, there is the phase in which the nature of the samples itself determines the word's extension; for the formula, to repeat, is that the extension of a word is determined by the environment itself, and that what aspect of the environment determines the extension is, in turn, determined by the nature of the samples Merlin pins his eye upon when using the word (whether or not anyone knows what that nature is).

Let us now spell, after Merlin, these two phases bit by bit, just to see whether it works. The word "water", say, fixes the extension {water} because we token — or better, enchant — "water" when standing in an indexing causal relation to samples of water. Crucially, we need not know what the nature of those samples is, so long as it is the same for all or most tokenings of the word under the conjuring circumstances. In other words, as regards the first phase of Merlin's formula, no contribution of environment to what we mean by "water" is yet involved. In the second phase, we call upon the natural kind of those samples — that is, the real essence <water> — to determine the identity of the extension {water}, ignorant though we may be as to the real constitution of <water>. Hence, *scarabhellcatphiltrebrew*, we have it that the environment, *water itself*, contributes to our meaning of "water"; so the meaning of "water" determines the extension {water}!

Or have we? To say, in the second phase, that the natural kind <water> fixes the identity of the extension {water}, is to say that the set of particulars which are water is defined by a certain natural kind, the property <water>; not that water as such contributes to what we *mean* by our token "water", nor that our meaning of "water" determines the extension {water}; which is to say, no contribution of real water to the meaning of the symbol "water" is involved; in general, the natural environment does not contribute to meaning. But then, Merlin's magic formula is a bluff! He has no thaumaturgic access to *de re* significance! Down with Merlin!

Quite apart from his formula, though, there is a simple way of seeing what lies behind Merlin's tricks: it is that although most people, after the Great Behavioural Revolution, have rejected the old oligarchic principle of extension-meaning supervenience, allowing themselves to be reduced to stimulus-response machines, the sly fox Merlin, whilst paying a lip service to popular ideology, has appropriated it for his ambitions. For his doctrine of the contribution of environment to meaning is but the principle of extension-meaning supervenience with a false moustache; it says really that the extension of a symbol either *is* or *is required for* its meaning. The people nowadays no longer believe the former option, the extensional theory of meaning; they would not confuse the Morning Star with the meaning of "Morning Star", or a dollar coin with the meaning of "dollar coin"; but the latter option still finds a rear wicket to smuggle the extensional theory back in disguise: if not as the whole account of meaning, at least as a part, joined with stereotypes or with overt verbal use, and called "the contribution of environment"; and, as Merlin knows, the latter option has just as potent thaumaturgic virtues as the former. There is only one way to break Merlin's power definitively, and that is to reject the extensional theory of meaning in every guise. The extensional theory is false not only *as a matter of fact*, it is *necessarily false*, false by virtue of what it says; it is not a genuine theory of meaning. The meanings of symbols must be one way or another aspects of the symbols, and so cannot be the particulars, or sets of particulars, or the real essences of particulars, which make up the non-symbolic natural environment.

The moral? Should you ever meet with Merlin again, hold your ground: the environment cannot fix the identity of meaning, or even contribute to it.

Chapter 3

Sentence-based Semantics

Early Steps toward Semantic Holism

3.1 Motivation for Semantic Holism

In modern Analytic Philosophy, semantic holism grew out of dissatisfaction with verificationist accounts of meaning of the logical positivists, and it is associated above all with W.V.O. Quine. But early steps toward semantic holism have been taken by the verificationists themselves, and even earlier by the founders of the Analytic tradition, Gottlob Frege and Bertrand Russell. In this chapter, we shall discuss these early moves toward holism, including the key semantical views of Frege, Russell, and Carnap, with special attention to the way their positions became reformulated in Quine's work.

Quine (1951; 1960) is concerned to show that, on the one hand, semantic *verificationism* of the logical positivists is false and leads to what Putnam and others (rather than Quine) call "meaning holism"; and, on the other hand, semantic *behaviourism* is false and likewise leads to holism. Verificationism and behaviourism are the only *reductive* semantical theories Quine seriously considers. In addition to these, Quine briefly contemplates, and rejects, three 'minor' reductionist alternatives: the theory that meanings are mind-independent universals; the theory that meanings are mental universals; and the extensional theory of meaning. The global argument for holism proceeds from the failure of semantic verificationism and behaviourism — together with the failure of the 'minor' alternatives — to the failure of semantic reductionism *in toto*: that is, to semantic *nihilism* insofar as the traditional notion of meaning is concerned; and it proceeds from the *way* reductionism fails to the conclusion that meaning — or *significance*, as Quine prefers to say — if it is anything, must be holistic (in a sense to be specified) and irreducible to non-semantic matters of fact. The argument begins as follows.

Quine (1948) speaks disparagingly of *Plato's beard*. This is the doctrine, according to Quine, that a term — whether singular or general — can be meaningful only if there is a unique entity which the term refers to: if not a particular object (for a singular term), then a mind-independent universal, a 'form in Platonic heaven' (for a general term). In other words, Plato's beard is the view that every meaning-bearing term must be a *name* for a certain entity. (As I have described it, Plato's beard differs slightly from Quine's original. Quine initially applies the doctrine to singular terms only; see (1948: 2–5, 9); then he goes on to extend its application to cover general terms as well (*ibid.*: 11). I have compounded these two aspects from the start.) Plato's beard is best regarded as a version of the Conservative rule of semantic individuation requiring that a difference in referent or extension should necessitate a difference in meaning, with mind-independent or *real* universals as referents of general terms, and concrete objects as referents of singular terms. More generally, Plato's beard may be taken as a cluster of semantical theories, including the extensional theory of meaning and also Fodor's referential account of mental content; in fact, any account based on the principle that the meaning of a symbol determines the symbol's referent — whether the referent be a real universal, or an extension comprising particulars instantiating the universal, or just a particular object or set of objects — may be regarded as a variant of Plato's beard. Though it is intended as a gist of Plato's position on meaning, we should bear in mind that Plato's beard is Quine's rendition of Plato, not Plato himself; it is arguable that Plato would have trimmed any outgrowth Quine imparts to him.

Quine dismisses Plato's beard as bad semantics and bad metaphysics. Very generally, his reason why it is bad semantics and bad metaphysics is two-fold: firstly, because it requires a *term-by-term* attribution of semantic properties; secondly, because it fails to *separate* the meaningfulness of a term from the term's reference, or determination of extension. Much of Chapter 4 will be spent on Quine's charges relating to the second problem with Plato's beard, and on the consequences he draws from it. In this chapter, we shall look closely into the first problem; and this will give us an occasion to trace the early moves Analytic Philosophy has made away from term-based semantics and toward semantic holism.

As Quine rightly points out, Plato's beard implies the traditional view that the primary bearers of meaning must be *terms* regarded one-by-one, in isolation from whatever linguistic contexts they occur in; and that sentences and other complex linguistic expressions must derive their significance from the meanings of their constituent terms. Why is this bad semantics? The reasons Quine (1948) finds persuasive against term-by-term semantic individuation have to do with ontological commitment. Consider the singular term "Pegasus". It follows from the term-by-term semantics of Plato's beard that Pegasus exists, granted only that "Pegasus" is meaningful. Suppose now that "Pegasus" is treated as a general term: that is, as the predicate "is Pegasus", or "pegasises". It follows that a universal — the property of being Pegasus, or (using my nomenclature) the real essence <pegasus> — exists. Quine finds both of these alternatives unacceptable: the former, that Pegasus exists, is plainly false; the latter, that a universal such as <pegasus> exists, is bad metaphysics. ('Metaphysics', for Quine, is just the business of telling what things there are, which would be better called "ontology"; Quinian 'metaphysics' is not to be conflated with the classical metaphysics of, say, Descartes, or the implicit metaphysics of Locke, let alone Plato.) Yet both of these consequences could be forestalled, Quine says, simply by abandoning term-by-term semantics; *i.e.*, by refusing to grant that "Pegasus" taken singly is meaningful.

Moreover, the abandonment of term-based semantic individuation need not be *ad hoc*, designed for the purposes of disentangling Plato's beard; this is because there are principled arguments to the conclusion that the primary bearers of meaning should be, not individual terms, but whole *sentences* (see Quine (1948: 5–8; 1951: 39)). These arguments are due to Frege and Russell, and have been used and transformed into a verificationist theory of meaning by Carnap; they are also the *terminus a quo* for Quine's holism. Quine regards the arguments as "an important reorientation in semantics — the reorientation whereby the primary vehicle of meaning came to be seen no longer in the term but in the statement" (1951: 39). We shall do well to look closely into these arguments, with a view to finding out how the first steps toward semantic holism were justified.

3.2 Frege on Term-based Semantics

Frege (1884): Suppose terms taken one by one, in isolation from whatever linguistic contexts they occur in, are bearers of semantic properties; and suppose one's problem is to specify, in general, what these properties consist in, granted only the Conservative principle that such properties must be the extension-determiners for the terms. Then one has no alternative but to adopt semantic *mentalism*: the view, according to Frege, that the property of the meaning of a term is the property of being associated with a mental representation which determines the extension of the term. More accurately, assuming the imagistic psychology Frege considers, the property of the meaning of an isolated term must be the property of being associated with a mental image — a *vorstellung*, in Frege's own nomenclature. But the mental image associated with a term (like Putnam's mental descriptions) may vary from individual to individual and from time to time, and need not comprise a sufficient condition for the membership in the extension of the term; that is, the mental image may not be the same or universal for every individual who knows the meaning of the term, and need not be an extension-determiner for the term. It follows that the mental image cannot be the meaning of the term; more explicitly, the property of the term's meaning cannot be the property of being associated with the image. Mentalism as a theory of meaning must be therefore false; and since term-based semantics requires mentalism, term-based semantics must be false; which is to say, terms cannot be bearers of semantic properties one by one, in isolation from whatever linguistic contexts they occur in (see Frege (1884: x, §§ 59–60)). Let us call this Frege's "tactical" argument against term-based semantics.

But why should *sentences*, if not terms, be the primary bearers of meaning? In order to answer this question, we must understand what may be regarded as Frege's *strategic* argument against term-based semantics, also due to Frege (1884), and developed in Frege (1893; 1903). The argument is as follows. Suppose term-based semantics is true and consider, in particular, terms referring to *numbers*. Since term-based semantics requires imagistic mentalism, we have it that for each number-referring term there must be an associated *mental image* which *determines* (is an image of) the *corresponding number*, or else the term is *meaningless*. Frege finds both of these alternatives unacceptable.

The *former* alternative is unacceptable for two closely related reasons. Firstly, it is unacceptable because — although every term, even a number-referring term, may be associated with a mental image — there can be no mental images of, specifically, numbers. A number "is not in fact either anything sensible or a property of an external thing" (Frege 1884: 70^{e}); it is therefore not the kind of thing which can be determined by an image; *i.e.*, it is not such that a difference in the number referred to by a term can depend on a difference in the mental image associated with the term. Secondly, the former alternative is unacceptable because it implies that *sentences* — or better, *statements*, taken as sentence tokens — of arithmetic are *synthetic* (*i.e.*, according to Frege, true by virtue of their conformation with empirical matters of fact) rather than *analytic* (*i.e.*, provable from sentential meaning in an axiomatic system, according to Frege); but arithmetical objects, such as numbers and relations among numbers, are 'not either anything sensible or a property of an external thing', so arithmetical statements cannot depend for their evaluation on empirical matters of fact, and hence cannot be synthetic. (See Frege (1884: §§ 5, 7–10, 12–14, 58–61) for further details of this argument, notably the strictures against Kant and Mill.)

The *latter* alternative is unacceptable because — unless arithmetic is but a formalistic 'game with pieces' (which it is not, see Frege (1903: § 93)) — arithmetical statements must have a domain of non-empirical application; for "it is applicability alone which elevates arithmetic from a game to the rank of science" (1903: 187). In other words, arithmetical statements require a non-empirical domain of interpretation and epistemic evaluation; hence numbers and relations of numbers must (be postulated to) exist, and number-referring terms cannot be meaningless, despite the impossibility of there being mental images of numbers, and so the impossibility of determining the extensions of number-referring terms in isolation, one by one. (See, *e.g.*, Frege (1903: §§ 89–93) for some of the arguments against formalism.)

But how could one be committed to the existence of numbers, assuming it is impossible to form *mental representations* which determine the extensions of number-referring terms? This is what may be properly called "Frege's problem". (See Frege (1884: § 62) for the problem; also Frege (1903: Appendix), especially the last paragraph.) Frege's solution is as follows.

To determine the extensions of number-referring terms, one must define the meaning (or 'fix the sense', according to the standard rendition of Frege's terminology) of *numerical identities*; that is, sentences such as "1+2 = 3"; further, one must fix the sense of the identities by expressing them as sentences of a certain kind of language: in effect, the language of pure second-order quantification logic with predicates for identity and set-membership (that is, the language of set theory). Why does this determine the extensions of number-referring terms? According to Frege, all statements of the logical language are *analytic* (either analytically true or analytically false); therefore the statements and, *in their contexts*, their constituent terms must have determinate semantic properties, hence determinate extensions; and since, as Frege purports to prove, every numerical identity is *synonymous* with some analytic statement of the logical language, it follows that the identities and their constituent terms — the terms referring to numbers and relations of numbers — must also have determinate semantic properties, and so determinate extensions; which is to say, numbers and relations of numbers must exist, granted that number-referring terms are bearers of semantic properties not one by one, but only in the contexts of sentences or statements, and assuming the analyticity of Frege's logical language. (See Frege (1884: §§ 62, 68–85; 1893: §§ 0–52; 1903) for the exposition and solution of his problem.) We shall now turn to see whether Frege's arguments hold out.

3.2.1 Imagistic mentalism: variability in *vorstellungen*.

Frege's *tactical argument* says that term-based semantics cannot be true since it depends on imagistic mentalism, the view that the meaning of a term consists in its being associated with a *vorstellung*, or mental image, which determines the term's extension, and since the image may vary from individual to individual and, for each individual, from time to time, and need not comprise a sufficient condition for the membership in the extension of the term. But though Frege is right in that term-based semantic individuation does depend on semantic mentalism, the mental representation associated with a public term need not and, in general, cannot be an image; this is just to note that term-based semantics need not rest on imagistic psychology. The only interesting aspect of Frege's failing here is that it exemplifies a kind of fallacy that has been repeated time and again throughout the Analytic movement: the philosopher would look at the current state of cognitive science, and take it for granted; thus Frege relied on 19th-century imagistic brand of empiricism, which had little to do with classical empiricism of, in particular, Locke; Quine relied on behaviourism in psychology and linguistics; Fodor on computational psychology; and so forth.

The tactical argument can be reformulated to accommodate the objection. Frege might say that term-based semantics is untenable, given its dependence on semantic mentalism, since psychological representations of any kind — *e.g.*, mental *descriptions* rather than images — may vary from individual to individual and from time to time, and need not comprise sufficient conditions for the membership in the extensions of their associated terms. This is the line of argument adopted by Putnam (1975; 1988), as in the elm-beech case, *etc.*

Frege's argument so reformulated fails nevertheless to rule out term-based semantics. One reason is that the Fodorian position, that all *prima-facie* lexical representations are simple, unstructured and unlearned, yet still determine their extensions, could be considered as true. But of course I am not defending Fodor. The main reason why Frege's tactical argument does not tell against term-based semantics is that, like Fodor's position, it rests on the Conservative rule for fixing the identity of meaning, that a difference in extension or referent necessitates a difference in meaning; that is, on the principle of extension-meaning supervenience, the same principle which is an undoing for Fodor. This wrong principle is a piece of misguided common sense, tangled in use-mention fallacies, and produced by a confused intuition that the meaning of a symbol either *is* the symbol's extension, or *determines* its extension. Classical term-based semantic mentalism — of Locke, Descartes, and many others — did not rely on this principle. The best source on this issue is Book III of Locke's *Essay*. Locke shows therein that although some ideas, such as ideas of natural numbers representing what he calls "simple modes", have a perfectly defined and universal semantic identity, most ideas we entertain are of substances or mixed modes, and these represent *nominal*, not *real essences*, which may and do vary from individual to individual or from time to time, and accordingly need not determine the class of particulars which may be said to instantiate the *real* essences. Thus, the idea **marigold** denotes a nominal essence [marigold], and [marigold] does not fix the extension {marigold}. Using my nomenclature, we may say that **marigold** *refers to* the real essence <marigold>, and hence to the particular flowers which are marigold; but this does not, nor is it required to, define the *meaning* of the idea, let alone the meaning of the word "marigold". The meaning of the word depends entirely on the semantic identity of the idea *in the mind of the speaker who uses the word on such and such an occasion*; as such, it may vary somewhat from what other speakers mean, or have in mind, when using the word, and it may vary from time to time as the same speaker's idea varies. But, what is important for our present concerns, none of this is any trouble for term-based individuation of meaning, whether in public language or in the mental code. The misconception that variations in ideas rule out term-based attribution of meaning is due to the wrong principle of extension-meaning supervenience; specifically, to the common failure to distinguish between what I have called "denoting" and "referring": words mean because the ideas in the mind of the speaker who uses them on such and such an occasion *denote* certain nominal properties; and they *refer* to their extensions, insofar as certain particulars may be said to partake of the properties; but referring does not fix the identity of meaning, only denoting does.

One might ask, what about number-referring terms: is it not the case that these terms and the corresponding ideas do determine their extensions or referents, so that their denotation and reference coincide? With the exception of Kant, the classical position does not prohibit that some complex ideas be both *clear and distinct* (well-defined and unique), *uniform* for all minds in a linguistic community, and *veridical* with respect to the environment. The complex idea **two** is, for Locke, just the idea of one added to one (where the ideas of unity and addition are semantically simple); it is the same for all speakers who know the meaning of the word "two"; and it is perfectly applicable to the world as it is. In such cases, Locke's view is that the essence represented by the idea is, in some sense, both nominal and real; to put it in my terminology, the denoting and referring of the idea **two**, and the public word "two", coincide. In general, it is a profound issue in the Classical Theory of Mind whether, if at all, the mind is capable of veridically representing the world as it is. We need not worry about it at present; all we need is to establish, as we have, that the charge of variability in mental representation does not impair term-based individuation of meaning; in particular, Frege's tactical argument does not undermine the project of term-based semantics.

3.2.2 Number-referring terms in axiomatic proofs.

Let us turn to Frege's strategic argument against term-based and for sentence-based semantics. It says that term-based semantics must be replaced with sentence-based semantics since the only way to prove that numbers and relations of numbers exist is to show that number-referring *terms* are meaningful; and the only way to show that number-referring terms are meaningful is to show that each *numerical identity* — a *statement* such as "$1+2 = 3$" — is semantically identical to some statement of the language of pure second-order quantification logic with predicates for identity and set-membership, and therefore analytic, or provable as true or false on the grounds of its meaning in an axiomatic system of that logic; hence it is taken to follow, since meaning determines extension, that numbers and relations of numbers do exist, though terms are meaningful only in the context of statements, not in isolation.

We should bear in mind that Frege's tactical argument against term-based semantics rules out for him the possibility that the semantics for the fundamental set-theoretic language be itself term-based. So Frege cannot say that, at bottom, the semantics for number terms is really term-based. But then, why should we accept Frege's assumption that his fundamental set-theoretic language is meaningful? Again, why should we accept his claim that statements of the language are analytic? Frege does not, in fact, provide anything like an explicit semantics for the language. Instead, he offers us an *axiomatic system* which he assumes to be semantically perspicuous; the system is intended to encompass all analytically true statements of the language; and analytic truth is assumed to consist in being provable from the axioms. The meaningfulness of the set-theoretic language as a whole, and hence also the meaningfulness of arithmetical terms and the existence of numbers, is then guaranteed provided each statement of the language is either analytically true or analytically false; that is, provable either as true or as false from the axioms. (This is why Russell's 'paradox' was such bad news for Frege; see Chapter 10.) Finally, since axiomatic proofs proceed *statement by statement* rather than *term by term*, it seems to follow that the primary bearers of meaning must be sentences, and that terms are bearers of meaning only in the context of a sentence.

There are at least two flaws in this reasoning. Firstly, axiomatic proofs — even proofs of such numerical identities as "1+2 = 3" — are used only by a small group of rather queer people, logicians and mathematicians; and by them only when they are earning their living. Vast majority of people, and most off-duty logicians, are able to prove and know with certainty that 1+2 = 3 without any axiomatic reasoning whatever, and many without so much as having heard of axiomatic systems and statement-by-statement proofs. They are able to do this because the mind, it is clear, has a way of knowing with certainty (a way of proof as yet unknown to formal logicians) which relies not on any axiomatic system but on *term-by-term analysis*; thus people know that 1+2 = 3 because they understand the *terms* "1", "2", "3", "+", and "=", and can sort out that the terms semantically agree in the proposition that 1+2 = 3. I will say that the mind proves such propositions *ex terminis*, from the constituent *terms* rather than from some other statements such as axioms; and will devote much of Chapters 7 and 9 to spelling out what *ex terminis* reasoning consists in. For the moment, let us note that even if Frege's argument for the existence of numbers were correct — consisting as it does in that numerical identities are provable in an axiomatic system and so *analytic*, and so *meaningful*, so that *in the context of the identities* number-referring terms must have *determinate extensions*, so that numbers must exist — this way of proof of the existence of numbers would be no evidence for the claim that sentences rather than terms are the primary bearers of meaning. For, on the one hand, the *way* of proof of the existence of numbers is not relevant to the issue of the fundamental bearers of meaning; Frege's way of proof runs *via* axiomatic reasoning, but other proofs might rely on *non-axiomatic* forms of reasoning.

On the other hand, supposing there were some original manner of reasoning to which all including axiomatic forms of reasoning were reducible, and supposing this original manner of reasoning were relevant to the issue of the fundamental bearers of meaning, this would be a small comfort to Frege. For such an original manner of reasoning would almost certainly *not* be axiomatic and statement-based; otherwise only formal logicians and mathematicians would be capable of knowing anything with certainty (if they could agree what the original axiomatic system is); common people would be lost adding one to two.

The second flaw in Frege's strategic argument against term-based and for sentence-based semantics is that the argument is *verificationist*; at any rate, it prepares the ground for semantic verificationism, since it makes meaningfulness dependent on an (axiomatic) *way of reasoning and confirmation*. It is then only a short step to the proposal that the meaning of a statement *consists in* its manner of confirmation. But verificationism we shall discuss in Section 3.4.

3.3 Russell on Term-based Semantics

Russell (1905): Suppose term-based semantics is correct and consider terms purporting to refer to objects which do not exist: "the present King of France", *etc.* Since term-based semantics requires that each meaningful term determine its extension or referent, it follows that the term "the present King of France" — and consequently the sentence "the present King of France is bald" — is meaningless, or else the present King of France exists. The former alternative is untenable since the sentence "the present King of France is bald" is false; hence, since it has a truth-value, it cannot be meaningless; and since terms rather than sentences are the primary bearers of meaning, "the present King of France" cannot be meaningless. The latter alternative is untenable since, according to it, the present King of France would both exist and not exist. In general, term-based semantics implies that terms purporting to refer to non-existent objects must be meaningless, or the objects must in some sense exist. The former option cannot be true since sentences involving such terms — and hence the terms themselves, given that terms rather than sentences are the primary bearers of meaning — are meaningful. The latter cannot be true since it implies that the objects such terms purport to represent both do and do not exist. So term-based semantics cannot be true. (See Russell (1905: 45–48) for this argument.)

But if so, how could one veridically assert that something does not exist, assuming it is impossible within the framework of term-based semantics to refer to non-existent objects? We shall regard this as *Russell's problem*. Russell's solution is that sentences, not terms, are the primary bearers of meaning, and terms are *meaningless even in the context of sentences* (contrary to Frege's position). His argument is as follows.

To assert of a non-existent object that it does not exist — to assert anything meaningful of a non-existent object — one must, to begin with, express the term purporting to refer to the object as what Russell calls "a denoting phrase". Denoting phrases are expressions such as "the author of *Waverley*", "the present King of France", and so forth. They are compound expressions consisting of a quantifier such as "the", "each", "some", "all", and a predicate such as "author of *Waverley*", "present King of France", *etc*. Some terms are denoting phrases as they stand; other terms, such as "Apollo", need to be expressed as denoting phrases by means of definitions: for example, "Apollo" becomes a denoting phrase when it is replaced with "what the classical dictionary tells us is meant by Apollo, say 'the sun-god'" (1905: 54).

Further, one must formulate certain semantical rules for interpreting sentences in which denoting phrases occur. The rules Russell puts forth amount to the standard rules for quantifier terms. The phrases "everything", "something", and "nothing" are taken as "the most primitive of denoting phrases" (1905: 42). The interpretation of *sentences* involving these phrases is to be determined in terms of the 'ultimate and indefinable' notion "$C(x)$ is always true", or "$C(x)$ is true for every value of the variable x", by the Fregean rules for logical quantifiers: *viz*., "everything is C" means the same as "$C(x)$ is true for every value of the variable x"; "something is C" means the same as "$C(x)$ is true for at least one value of the variable x"; and "nothing is C" means "$C(x)$ is true for no value of the variable x". In addition to these, Russell puts forth a derived rule for the interpretation of sentences involving the so-called "definite descriptions", which are denoting phrases made of the definite article "the"; *viz*., "the F is G" means the same as "there is a value of the variable x such that, for that value of x and for every value of the variable y such that $F(y)$ is true, both $F(x)$ and $G(x)$ and $y=x$ are true". (This is commonly called "Russell's theory of definite descriptions"; yet, like the foregoing rules for 'the most primitive of denoting phrases', the rule is really due to Frege; see, for instance, Frege (1884: § 55).) According to Russell, these are all the rules necessary for the solution of his problem. How do the rules work to solve it?

Recall that denoting phrases or terms cannot be bearers of meaning one by one, in isolation from their sentential contexts: "... a denoting phrase is essentially *part* of a sentence, and does not, like most single words, have any significance on its own account" (*ibid*.: 51); again, "... denoting phrases never have any meaning in themselves, but ... every proposition in whose verbal expression they occur has a meaning" (*ibid*.: 43). So far as concerns rejecting term-based and adopting sentence-based semantics, Russell follows Frege. But Frege's problem is to explain how certain entities, numbers, can be said to exist, given that the terms referring to them cannot be bearers of extension-determining semantic properties one by one; and, to solve his problem, Frege proposes to show that terms referring to numbers are bearers of such properties in the context of certain analytic statements, so that numbers can be said to exist. In contrast, Russell's problem is to explain how certain alleged entities can be said not to exist, and in general how it is possible to assert anything meaningful about non-existent objects. To solve his problem, Russell proposes to show, by means of his semantical rules for sentences involving denoting phrases, that all denoting phrases are meaningless even in the context meaningful sentences, so that the alleged entities can be said not to exist.

He argues thus. Take, for example, the denoting phrase "a man" in the context of the sentence "I met a man". This sentence means that "I met x and x is human" is true for at least one value of the variable x; according to Russell, "[t]his leaves 'a man', by itself, wholly destitute of meaning, but gives a meaning to every proposition in whose verbal expression 'a man' occurs" (1905: 43). Analogously, consider the denoting phrase "the author of *Waverley*" in the context of the sentence "the author of *Waverley* was a man". The sentence asserts that "x wrote *Waverley*" is true for one and only one value of the variable x, and "x was a man" is true for that value of x. Again, the analysis assigns a meaning to the sentence, but leaves the phrase "the author of *Waverley*", as Russell puts it, 'wholly destitute of meaning': "the phrase *per se* has no meaning, because in any proposition in which it occurs the proposition, fully expressed, does not contain the phrase, which has been broken up" (1905: 51). Generally, the set of rules for interpreting sentences "gives a reduction of all propositions in which denoting phrases occur to forms in which no such phrases occur" (*ibid.*: 45); so all denoting phrases, whatever they purport to refer to, are meaningless even in the context of meaningful sentences. In this manner, we may speak meaningfully of Apollo, the present King of France, the island than which none greater can be conceived, or any non-existent thing that will please our fancy, provided only we express ourselves in grammatical sentences.

But if *all* denoting phrases, not just those purporting to refer to non-existent objects, are meaningless even in the context of sentences, how could any phrase — such as "the author of *Waverley*" — refer (or denote, in Russell's terminology)? Russell's account of denotation (*i.e.*, my reference) amounts to saying that, to determine the extension of a term, one must express the term as a *definite description*, and one must find and interpret, by means of his rule for definite descriptions, a *true* statement of *identity* involving the definite description; the extension of the term is then whatever that statement says that it is (see (1905: 51)). Notice that this 'explanation of denotation' is akin to Frege's account of reference for number-terms, in that both Russell and Frege demand that, to determine the extensions of terms, one must fix the sense of certain statements — *viz.*, statements of identity — and one must fix the sense of the statements by expressing them in a certain canonical logical language; the accounts differ, however, in that Frege requires the canonical statements to be analytic, whereas Russell puts up with true contingent identities. Let us now look at Russell's arguments critically.

3.3.1 Meaningful terms referring to non-existent objects.

Russell's argument against term-by-term semantics is that, supposing it correct, terms such as "the present King of France" would have to be meaningless, else their referents would in some sense have to exist, neither option being acceptable given our assumptions. The underlying semantical intuition is that, as regards term-based semantics, each meaning-bearing term must represent a certain unique entity; and this unique entity is taken to be the referent or extension of the term. I concede that term-based semantics requires that for every meaningful term there must be an entity the term represents, but deny that the entity must be a particular object such as the present King of France. The meaning of the term "the present King of France" may consist — I think it does — in its denoting the *nominal property* [the present King of France], the identity of which is fixed by some mental description such as **the present King of France**. Granted the term is so meaningful, it clearly does not follow that the present King of France both does and does not exist; what follows is that no extant entity partakes of the property [the present King of France]. Similarly, *mutatis mutandis*, for other terms denoting non-existent objects.

But, one might say, the intuitive requirement for term-based attribution of meaning is that for each meaningful term there must be a *unique* entity the term represents; and the mental descriptions determining the identity of the nominal property [the present King of France] may vary from individual to individual and, for each individual, from time to time. So what is *the* meaning of "the present King of France"? To begin with, whenever the term is tokened or used, it cannot express any idea but that in the mind of the speaker who uses the term on that occasion, and hence cannot denote any nominal property but that fixed by the speaker's idea at that time. So the term may indeed vary in meaning from speaker to speaker, or from one occasion of tokening to another. Nevertheless, we can still attribute a socially uniform meaning to the term, insofar as the mental descriptions in the minds of the individual speakers of a linguistic community tend toward sameness, under social pressures, as toward a limit. So long as the descriptions associated with the same public term in the minds of individual speakers are similar, communication can proceed well; yet such social uniformity does not change the fact that meaning is fundamentally a personal matter.

3.3.2 The sentence-based semantics of denoting phrases.

Term-based semantics has no difficulty with terms purporting to refer to non-existent objects. But how about Russell's contention that the primary bearers of meaning are sentences rather than terms? His sentence-based semantics is supposed to resolve the problem of asserting of a non-existent object that it does not exist; or, in general, the problem of asserting anything meaningful of a non-existent object. We have already seen what the correct answer should be: meaningful discourse depends on denoting nominal properties, not on referring to particulars, whether existent or non-existent. Still, it will be worthwhile to look into Russell's semantics, pointing out a few among the many errors in it.

Firstly, Russell says that all terms, including terms allegedly referring to non-existent objects, must be expressed and regarded as denoting phrases: *i.e.*, compound expressions such as "the present King of France", "a man", *etc.*, consisting of a quantifier such as "the", "a", "all", and a predicate such as "man" or "present King of France". This requirement is entirely *ad hoc*, tailored solely for the purpose of Russell's 'solution' concerning meaningful discourse about non-existent objects, and without an independent justification. The gist of it is that every term must be conjoined with one or another quantifier term; and the sole rationale for this is to use the Fregean contextual rules for the quantifiers, claiming that since these rules are specified relative to the contexts of sentences, it follows that semantics must be sentence-based.

Secondly, Russell takes it that the Fregean rules for quantifier terms are semantical rules for *interpreting sentences* which involve the denoting phrases. But the rules are simply rules for the quantifiers; they are rules that help to work out the *truth-conditions*, rather than the meanings or interpretations, of sentences involving the quantifiers; and, as such, they are given in sentential contexts. This does not show that sentences rather than terms are the primary bearers of meaning. Further, since the truth-conditional theory of meaning — being a species of the extensional theory — cannot be correct, the fact that the rules for quantifiers are specified in sentential contexts does not even show that the semantics of the quantifier terms themselves must be sentence-based; all it shows is that the specification of *truth-conditions* for statements involving quantifiers must be contextual, or sentence-based.

Thirdly, Russell uses the Fregean rules for quantifiers to argue that denoting phrases such as "a man" are meaningless, whether in isolation or in the context of meaningful sentences such as "I met a man", since "I met a man" means that 'I met x and x is human' is true for at least one value of the variable x, and this, Russell says, "leaves 'a man', by itself, wholly destitute of meaning" (*op. cit.*). But here Russell interprets nothing at all: he applies the rule for the quantifier term "a", but leaves the rest of the sentence intact, taking no notice whatever of the terms "man", "I", and "met".

Lastly, Russell holds that although all terms or denoting phrases are meaningless both in isolation and in the context of meaningful sentences, some do veridically represent certain objects; that is, *terms may refer despite* *being meaningless*. His 'explanation of denotation' (*i.e.*, reference) is that a term refers just in case, expressing the term as a definite description, there is a *true* statement of *identity* involving the definite description; and the referent of the term is whatever that statement says that it is. But this is not a genuine account of reference. For one thing, it rests on the notion of truth, and that in turn presupposes the notion of reference; Russell's 'explanation of denotation' simply begs the question. But more seriously, what we expect to learn from an account of reference is how symbols can be related to particular objects, and real aspects of particular objects, in the environment. Our problem is not what the *term* "reference" *means*; for that is a question about the symbol "reference", an entirely different matter. We wish to know what the relation of reference between a symbol and certain particulars or their real properties consists in: what is the *nature* of reference. Russell's 'explanation of denotation' does not even begin to address the issue of the relation between, say, "the author of *Waverley*" and Walter Scott.

It is very plain that both Frege's and Russell's arguments against term-based and for sentence-based individuation of meaning fail. Yet these arguments are the sole basis for the subsequent verificationist tenet that semantics must be sentence-based, as we shall see in the next section; and, in turn, the failure of semantic verificationism is the main, though not sole, basis for subsequent Quine's contentions that meaning is holistic rather than either term-based or sentence-based. However befuddled the Frege-Russell arguments against term-based and for sentence-based semantics may be, their importance for the history of Analytic Philosophy can be hardly over-estimated. Writers of textbook Analytic Philosophy regarded especially Russell's part as 'epoch making' (*e.g.*, Richards (1978: 72)); others took it for an exemplary case of philosophical analysis; still others for "a milestone in the development of contemporary philosophy, revealing once more Russell's inventiveness and striking originality in thought" (Marsh in Russell (1956: 39)). Marsh reports that another pillar of Analytic Philosophy, Professor Moore, has discovered an error in Russell's article, consisting in that the statement "Scott is the author of *Waverley*" means not the same as "Scott wrote *Waverley*", on account of an ambiguity in the verb "to write"; but Russell passed over it with equanimity ...

3.4 Carnap's Semantic Verificationism

Verificationism rests on the supposition that Frege and Russell have demonstrated conclusively that term-based semantics is untenable, and that the fundamental bearers of meaning must be sentences (see, for example, Carnap (1936–37: 2)). As regards the nature of sentential meaning, verificationism offers a fusion of problems of meaning with problems of knowledge:

> Two chief problems of the theory of knowledge are the question of meaning and the question of verification. The first question asks under what conditions a sentence has meaning... The second one asks how we get to know something, how we can find out whether a given sentence is true or false. The second question presupposes the first one. Obviously we must understand a sentence, i.e. we must know its meaning, before we can try to find out whether it is true or not. But ... there is a still closer connection between the two problems. In a certain sense, there is only one answer to the two questions. If we knew what it would be for a given sentence to be found true then we would know what its meaning is. And if for two sentences the conditions under which we would have to take them as true are the same, then they have the same meaning. Thus the meaning of a sentence is in a certain sense identical with the way we determine its truth or falsehood; and a sentence has meaning only if such a determination is possible. (Carnap 1936–37: 420)

In other words, a sentence is meaningful only if it is verifiable, and the meaning of the sentence is the method of the sentence's verification. What does the method of verification consist in? Carnap's theory of verification derives from a certain account of language originating in Russell (1918–19) and Wittgenstein (1922). According to that account, every meaningful sentence is, in principle, expressible as a logical construction from *atomic sentences* representing *atomic facts*.

Carnap (1928) assumes that atomic sentences are sentences such as "the sensation *red* is at space-time point (x, y, z, t)"; and that atomic facts are, correspondingly, sense-datum facts. Carnap (1932–33; 1936–37) assumes that atomic sentences represent publicly observable, physical states of affairs or physical events. The position of Carnap (1928) is, however, not that phenomenalism as opposed to materialism is true; accordingly, the position of Carnap (1932–33; 1936–37) is not that materialism is true. Carnap emphasises that these metaphysical theses are, even in principle, unverifiable and thus *meaningless*. In general, verificationism holds that empiricist theories of meaning — that is, theories which take themselves to hark back more or less to Locke's account of meaning and knowledge — must be metaphysically non-committal. The main historical reason for this is that Berkeley and Hume are alleged to have shown, respectively, that Locke's account of the acquisition of ideas allows neither of the acquisition of an idea of *physical* substance, nor of the acquisition of an idea of *mental* substance, so that the terms "physical substance" and "mental substance" must be, so far as empiricism is concerned, devoid of meaning. According to Carnap, one cannot therefore meaningfully choose between the metaphysical doctrines of materialism and phenomenalism; but one has a pragmatic choice between the associated scientific languages: one can choose whatever language, physicalistic or phenomenalistic, suits best for the purposes of science (see Carnap (1936–37: 427–31) for details).

A simple consequence of the Russell-Wittgenstein view of language is that a sentence cannot be meaningful unless it is, in principle, verifiable by a certain determinate method: the method of expressing the sentence as a logical construction from its atomic constituents, evaluating the atomic constituents for truth or falsehood with respect to the atomic facts, and computing the sentence's truth-value from the values of its constituents by the logical rules involved in its construction. The theory of verification adopted by Carnap (1928; 1932–33) closely follows this sort of account. However, Carnap (1936–37) finds it necessary to modify that account in two respects, which are important from the point of view of our history of how Analytic Philosophy gradually moved, *via* sentence-based semantics, toward semantic holism.

Firstly, he amends the implicit supposition that a meaningful sentence is *conclusively* verifiable by the aforesaid determinate method. His reason is that *universal sentences* — *e.g.*, sentences expressing natural laws — although meaningful, are nevertheless not conclusively verifiable by any such method; at best, they are confirmable to some degree of probability. A further reason is that *particular sentences* — those representing particular states of affairs or events — are also not conclusively verifiable, since the number of sentences or predictions which one could *infer from* a sentence such as "there is a white sheet of paper on this table", and which one would have to confirm in order to have a complete verification of the sentence, is infinite. (See Carnap (1936–37: 425) for this argument.)

Secondly, Carnap amends the supposition that every meaningful sentence is verifiable by a *determinate* method. His reason here is that since the verification of a sentence depends on the confirmation of other sentences or predictions inferred from it, and since the number of these other sentences may be indefinitely large, the verification of the sentence must be partly a matter of *pragmatic* decision; there must always be a conventional component in the verification of a (synthetic, *i.e.*, contingent) sentence, due to the fact that one must put a limit on the number of predictions to be inferred and confirmed in order to verify the sentence; and this limit is to some extent arbitrary, set by exigencies of research, *etc.*

In summary, then, Carnap's account of meaning is that a sentence is meaningful only if it is, at least in principle, verifiable; and the meaning of the sentence consists in the method of the sentence's verification. The method of verification includes computing the truth-value of the sentence from the truth-values of certain atomic observational sentences representing atomic facts. But, in general, the verification of any contingent sentence is *under-determined* by the observational sentences. This is because, firstly, the verification of not only universal but also particular sentences (and perhaps even of the atomic observational sentences) depends on the confirmation of an indefinite number of other sentences or predictions; and, secondly, because the problem of selecting which of these other sentences, and how many, are to bear on the verification has at best a pragmatic or conventional solution.

Verificationism has been much discussed over the past decades. Here I will point out three problems with it, which I think are not yet familiar in the literature. These will concern the three basic aspects of verificationist semantics: *(i)* that the primary bearers of meaning are sentences; *(ii)* that the meaning of a sentence is the method of verification of the sentence; *(iii)* that the verification of a sentence requires confirming indefinitely many other sentences or predictions inferable from it, and hence is always underdetermined by available evidence.

3.4.1 Sentences as fundamental bearers of meaning.

Verificationism rests on the supposition that Frege and Russell have shown that term-based semantics is untenable, and that the fundamental bearers of semantic properties are sentences. We have seen earlier that Frege and Russell failed either to rule out term-based semantics, or to justify sentence-based semantics.

3.4.2 Meaning as method of verification.

Verificationism claims that the meaning of a sentence consists in the method of confirmation of the sentence. There are a number of objections to this position: for example, that a statement may be and typically is confirmable by a variety of different methods without being ambiguous. Here is another argument. In the foregoing quotation, Carnap says that 'two chief problems of the theory of knowledge are the question of meaning and the question of verification', thus conflating semantics with epistemology. We shall see anon that this turns out to have untoward consequences. He goes on to say that 'the first question asks under what conditions a sentence has meaning, whereas the second asks how we get to know something, how we can find out whether a given sentence is true or false'; and that 'the second question presupposes the first one, since obviously we must understand a sentence — *i.e.*, we must know its meaning — before we can try to find out whether it is true or not'. Carnap now conflates the problem of determining the conditions for the *meaningfulness* of a sentence with that of determining the sentence's *meaning*; for he introduces the question of meaning as one which 'asks under what conditions a sentence has meaning', yet proceeds to say that this question is presupposed by the question of verification, in that 'obviously we must understand a sentence — *i.e.*, we must know its meaning — before we can try to find out whether it is true or not'. But the question of meaning must surely ask a great deal more than 'under what conditions a sentence has meaning'. So far, nothing too serious. The main difficulty begins when Carnap says that 'there is a still closer connection between the problem of meaning(fulness) and the problem of verification', and that 'there is only one answer to the two questions...'; that is, when he puts forth the verificationist account of meaning. For since, as he himself correctly admits, 'we must understand a sentence — *i.e.*, we must know its meaning — before we can try to find out whether it is true or not', there cannot be 'only one answer to the two questions'; otherwise we would have to know the method of verification of the sentence before we could address the question of its method of verification; correspondingly, we would have to know the meaning of the sentence before knowing the sentence's meaning.

3.4.3 Contingent verification as confirmation of indefinitely many predictions.

Carnap's (1936–37) view that the verification of any contingent sentence is under-determined by available evidence later became a major tenet of Quine's epistemic and semantic holism, which we shall deal with in much detail in Chapters 4–5. Here I will mention a confusion in Carnap's notions of evidence and inference. Carnap begins by noting that universal sentences — sentences purporting to represent natural laws — are not conclusively verifiable since the number of instances of such laws is infinite; he goes on to claim that "there is no fundamental difference between a universal sentence and a particular sentence with regard to verifiability but only a difference in degree" (1936–37: 425). To show this, he considers the particular sentence "there is a white sheet of paper on this table"; he says that "to ascertain whether this thing is paper, we may make a set of simple observations and then, if there still remains some doubt, we may make some physical and chemical experiments" (*ibid.*), aiming to verify further sentences or predictions which *follow from* the sentence. He argues that the "number of such predictions which we can derive from the sentence given is infinite; and therefore the sentence can never be completely verified" (*ibid.*).

But this is rather like arguing that when I see the lights turn green on an intersection, I must seek further evidence from the next driver (*and* the next, *and* the next), *and* ring the department of transport, *and* conduct physical experiments, *etc.*, to make *sure* I may proceed. The point is that doing all or some of this madness would indeed provide me with further evidence that the lights are green and I may proceed, but none of it is *necessary*; the evidence I have from *seeing* the lights turn green is not all the evidence I may get, but it is *sufficient* under normal circumstances. Analogously, Carnap presupposes that the particular statement "there is a white sheet of paper on this table" would be conclusively verified only if one were to verify *all* predictions inferable from it. On some accounts of inference, mathematical truths follow from any statement; that would indicate that Carnap's statement could not be verified before having proved all mathematical propositions. In general, since ($\alpha \vee \beta$) follows from α, no proposition α would be verifiable unless one were omniscient! Behind the commotion is Carnap's confused notion of evidence: he thinks that the evidence for a contingent statement can come only from what the statement implies, what is inferable from it. This is often so; but very often, evidence is what *implies* or *is sufficient for* the statement. To run an account of contingent verification which rules out the possibility of sufficient evidence only because, as a matter of fact, we can never exhaust all conceivable evidence, is to make such verification impossible *by fiat*.

We have seen that the key ingredients of semantic-*cum*-epistemic holism can be found in Carnap's verificationist account of meaning. In the next chapter, we shall turn to semantic holism proper. But it is already clear from our work so far that the early steps toward holism, which Quine hailed as 'an important reorientation in semantics — the reorientation whereby the primary vehicle of meaning came to be seen no longer in the term but in the statement', were taken with much titubation, and very likely in pursuit of a will-o'-the-wisp.

Chapter 4

Radical Empiricism I

4.1 Meaning and Non-existent Entities

I pointed out in Section 3.1 that Quine (1948) disparages the term-based semantics of Plato's beard on the grounds that some *prima facie* meaningful terms have no determinate extensions; that, for example, the term "Pegasus" is meaningful although Pegasus does not exist. Quine's problem is thus identical to Russell's problem: given that term-based semantics fails to account for meaningful discourse about non-existent objects and therefore must be false, how does one veridically assert that something does not exist; more generally, how does one assert anything meaningful of a non-existent object or class of objects? Quine's solution to the problem radically differs from Russell's solution. On the one hand, Quine forgoes the Russellian claim that terms are meaningless both in isolation and in context; that is, he allows of the meaningfulness of terms in larger linguistic contexts. On the other hand, however, he needs to avoid, so as to resolve Russell's problem, the Fregean conclusion that, in the context of larger linguistic structures, terms are still bearers of the Conservative extension-determining semantic properties. Quine consequently opts for what might be regarded as a *Wittgensteinian* solution to the issue: like Wittgenstein, he proposes to change radically the notion of meaning as extension-determiner. (*Cf.* Wittgenstein (1953: § 55) for a version of Russell's problem, and (*ibid.*: §§ 65–71, 75–84) for his resolution.) This chapter and the next will be devoted to showing just what Quine's alternative notion of meaning is, and what is wrong with it. I will divide the subject into four portions, to be spread between the two chapters. The present chapter will be concerned with his views on the relationship between meaning and reference, and on mental and mind-independent semantic universals; the next with his views on semantic verificationism and behaviourism. In each case, I will show how these positions are supposed to ensue in semantic holism, and why as matter of fact they do not. The last section of Chapter 5 will critically inspect Putnam's transformation of Quine's semantic holism into his own *pragmatic* or *internal realism.*

4.2 The Separation of Meaning and Reference

I mentioned in the last chapter that Quine dismisses Plato's beard as bad semantics and bad metaphysics; and that, on the one hand, Plato's beard is bad semantics, according to Quine, because it requires term-by-term attribution of semantic properties. On the other hand, Plato's beard is bad semantics because it fails to separate the meaningfulness of terms from the terms' determination of extension, or reference. Why is this bad semantics? In other words, what reasons are there for abandoning the Conservative notion of the meaning of a term as that which fixes the extension of the term? To answer this question, we shall look into Quine's views regarding three closely related versions of Plato's beard: *viz.*, the extensional theory of meaning; the theory that meanings are mind-independent semantic universals; and the theory that they are mental semantic universals. These theories are, for Quine, the minor reductionist alternatives (the major being verificationism and behaviourism). Briefly, Quine argues that each of these versions of Plato's beard, when spelt out more explicitly, is bad metaphysics, and concludes that — given the reasons why Plato's beard is bad metaphysics — the correct way to resolve such semantical puzzles as Russell's problem is to dissociate the meaningfulness of terms from the terms' determination of extension; *i.e.*, to abandon the Conservative notion of meaning as extension-determiner.

Consider, to begin with, the simplest version of Plato's beard: namely, the theory that the meaning of a term is the object, or set of objects, the term refers to. This is the so-called "extensional theory of meaning" (*cf.* Putnam (1975: 217)). According to Quine, the extensional theory is untenable because, as we have seen, some meaningful terms have empty extensions (*e.g.*, "Pegasus" is meaningful even though Pegasus does not exist); further, it is untenable since some pairs of non-synonymous terms have identical extensions: the terms "Morning Star" and "Evening Star" refer to the same object yet differ in meaning; again, the terms "creature with a heart" and "creature with a kidney" have identical extensions but different meaning. (See Quine (1948: 9; 1951: 21) for these arguments; the Morning-Star/Evening-Star example is due to Frege (1892).) The extensional theory must therefore be false.

Reflect now on the following emendation one might suggest in order to reclaim the extensional theory despite the preceding arguments. One might treat, in particular, the term "Pegasus" as a general rather than singular term, as the predicate "is Pegasus" or "pegasises"; and one might say that "Pegasus" refers to a universal; that is, the property <is Pegasus> or <pegasises>, thereby resolving the problem of terms with empty extensions. Analogously, one might say that the terms "creature with a heart" and "creature with a kidney" refer to, respectively, the properties <creature with a heart> and <creature with a kidney>; that "Morning

Star" refers to <Morning Star>, "Evening Star" to <Evening Star>, *etc.* This would seem to resolve the problem of co-extensive yet non-synonymous terms.

According to Quine, though, the extensional theory of meaning thus emended is bad metaphysics since it presupposes the existence of universals, namely, *properties*. Why is this bad metaphysics? Quine says that questions of ontology — questions about what properties, objects, or events there are (for him, metaphysics and ontology deal with the same problems) — cannot be decided absolutely and objectively, independently of the *conceptual scheme* one deploys to interpret one's experiences. Take, firstly, a *realistic* conceptual scheme based on the semantics of Plato's beard. The gist of such a scheme is that whenever there are entities (objects, events) which can be *said*, *veridically*, to have something in common, there must *be* some *property* the entities (objects, events) have in common — granted only the predicate purporting to refer to the commonality is *meaningful*. For example, whenever one can speak veridically of red houses, red roses, red sunsets, *etc.*, there must be something the houses, roses, sunsets, *etc.*, share in common — *viz.*, the property <red> — granted the predicate "red" is meaningful. Secondly, take a *nominalistic* conceptual scheme based on the alternative notion of meaning Quine proposes. The gist of this scheme is that one can speak, *veridically*, of a commonality shared by diverse entities (objects, events), yet deny that there is a property the entities (objects, events) have in common, although the predicate purporting to refer to the alleged commonality is meaningful (see Quine (1948: 10)). The two conceptual schemes are clearly incompatible; yet it is impossible to decide, Quine holds, which of these schemes — if any — is true of the world. Hence it is impossible to decide absolutely and objectively what properties (objects, events) there exist, or even whether or not there are such things as properties or universals. The postulation of absolute and objective universals is therefore bad metaphysics.

Further, the conceptual scheme based on Plato's beard is bad metaphysics because it is possible to adjudicate between the two schemes on the grounds of various *pragmatic* criteria concerning the schemes' simplicity and explanatory power; and because, with respect to these pragmatic criteria, the nominalistic scheme is superior to the realistic scheme of Plato's beard. The nominalistic conceptual scheme is *simpler*, more economical, in that it presupposes there being fewer (if any) abstract entities; specifically, it presupposes there being no such entities as properties (*e.g.*, no entity corresponding to the term "red"). In addition, nominalism gains ontological simplicity without losing any of its *explanatory power*; according to Quine, the realist is "no better off, in point of real explanatory power, for all the occult entities which he posits under such names as 'redness'" (1948: 10). The reason is that whenever the realist claims that some *particular* entity is red (is a house, rose, whatever), the nominalist may agree; the disagreement will set in only if the realist claims to refer to an abstract entity, an entity one cannot have, in principle, any experiences of. Therefore, given that experiences are the *explananda* of conceptual schemes (see Quine (1948: 10, 16)), nominalism suffers no loss of explanatory power in abandoning all reference to properties.

Finally, the nominalistic scheme is not only equal but superior to the realistic scheme, in that it comports better with — what Quine, rather than the realist, takes to be — empirical data concerning the *actual use of language*, and actual practices of positing and individuating properties in a conceptual scheme. This is best argued for in Quine (1960; 1969). The claim, more explicitly, is that actual linguistic use and actual practices of positing and individuating properties defy the realistic myth of *in pluribus unum* (or *the one in the many*); that is, the supposition that diverse entities gathered under the same general term (red roses, red houses, red sunsets, and so on) must share something in common (the property <red>), provided only the general term is meaningful. The argument, in an outline, is as follows.

Quine takes it as axiomatic that this inference holds:

> Language is a social art. In acquiring it we have to depend entirely on intersubjectively available cues as to what to say and when. Hence there is no justification for collating linguistic meanings, unless in terms of men's dispositions to respond overtly to socially observable stimulations. (1960: ix) (*Cf.* Quine (1960: §§ 1-2; 1969b: 26-27.)

Yet the 'intersubjectively available cues' and 'socially observable stimulations' necessary for the acquisition and correct use of a word or sentence retain, as Quine puts it, a 'subjective twist', in that they need not — indeed, cannot — be entirely identical for different users of the word or sentence. For example, consider words such as "square" or "red". The cues and stimulations required for the acquisition and use of "square" are not identical for different users of the word, since in any learning situation, such as when focussing on a square tile, the teacher and learner — in general, different users of the word — experience different scalene projections of the tile. Similarly, although "red" has less of a subjective twist than "square", still there are some differences, in any learning situation centred on a red object, between the stimulations and cues experienced by the teachers and learners involved, "insofar as reflections from the environment cause the red object to cast somewhat different tints to different points of view" (1960: 8). Problem: if so, what justification can there be for 'collating the meaning of a word in terms of dispositions to use the word in response to certain socially observable stimulations', granted the stimulations cannot be identical for different users of the word?

Quine puts forth, to answer this question, the doctrine of *e pluribus unum*, or *the one from the many*. This is the view that there is an objective pull which works to *regiment* the use of words such as "square", "red", *etc.*, on the basis of *resemblance* rather than identity of the cues and stimulations required for the acquisition and correct use of the words (where the standards of resemblance are constrained, on the one hand, by an innate 'quality space' or similarity metric — see Quine (1960: § 17; 1969c: 127–128) — and, on the other hand, by social training inasmuch as the regimentation of linguistic use is a result of social interface between learners and teachers; see Quine (1960: § 2)).

Consider now the effect the doctrine of *e pluribus unum* has on the realistic myth of *in pluribus unum*; that is, on the supposition that whenever there are entities which can be said, veridically, to have something in common, there must be some property they share in common, provided the term purporting to refer to the alleged commonality is meaningful. Since a general term subsumes experiences, and hence the objects or events occasioning the experiences, on the basis of *similarity* rather than identity in the relevant respect, it follows that the term can be meaningful — *i.e.*, according to the present proposal, can be used and applied correctly on occasions of the appropriate experiences — even when the experiences and objects (events) occasioning the experiences are not bound by any common attribute or property. The meaningfulness of general terms is therefore not dependent on the existence of properties. Despite this, Quine admits, common-sense realism does assume that general terms refer to general attributes of objects, or properties. Why so? Because talk of reference concerns first and foremost terms standing for particulars, and there it is justified enough; but people then extend referential talk to general terms, simply because general terms are grammatically analogous to particular terms, and unwittingly wind up as believers in the myth of *in pluribus unum*. (See Quine (1969a: 19, 13–16).)

Thus actual linguistic use and actual practices of positing and individuating properties, as described by the doctrine of *e pluribus unum*, defy the myth of *in pluribus unum*, the belief that when there are diverse particulars which can be veridically and meaningfully said to have something in common, there must be some property they share in common. The nominalistic conceptual scheme is therefore superior in explanatory power to the realistic scheme based on Plato's beard, and so Plato's beard — in particular, the emended version of the extensional theory of meaning — is bad metaphysics.

This argument is taken to show that semantics ought to dissociate the meaningfulness of terms from the terms' determination of extension, and so to abandon the Conservative notion of meaning as extension-determiner. It is taken to show that because, so long as a general term subsumes experiences, and the objects (events) occasioning the experiences, on the basis of resemblance rather than identity in the relevant respect, the meaningfulness of the term need not (indeed, cannot) depend on the term's determination of extension, for the term need not have a determinate extension; it can depend only, according to Quine, on the term's use and application *vis-à-vis* sundry similarity-bound experiences and objects (events) occasioning the experiences. That is to say, so long as a general term subsumes experiences and the objects (events) occasioning them on the basis of resemblance rather than identity in some respect, the extension of the term cannot supervene on the term's meaning; in other words, the term's meaning cannot be its extension-determiner. It follows that, in general, meaning and reference must be separated, and the common-sense notion of the meaning of a term as that which fixes the term's extension must be abandoned.

Lastly, we should note that Quine's alternative notion of meaning allows of an answer to Russell's problem. The problem, to repeat, is this: given that term-based semantics fails to account for meaningful discourse about non-existent objects and therefore must be false, how does one veridically assert that something does not exist, or anything whatever about a non-existent object? For Quine, the problem is resolved since the meaningfulness of a term is not dependent on the existence of a determinate extension of the term. Rather, it is dependent on a bunch of kindred experiences — we might say, a fuzzy set of experiences — stimulating the term's correct use and application; but the experiences need not have anything in common; in particular, they need not be occasioned by any one object or class of objects, nor by the instantiation of a property. So "Pegasus" can be meaningful although Pegasus does not exist. (Compare Quine's views with those of Wittgenstein (1953). For example, see (1953: § 1) for an exposition of the extensional theory, and also for the first objections against it; Wittgenstein, like Quine, contemplates an alternative 'language game', or conceptual scheme, grounded in the semantics of verbal behaviour rather than reference, and subsequently argues that actual linguistic use comports better with the alternative than the referential 'game'. See also (1953: §§ 66–71, *passim*) for Wittgenstein's arguments against the realistic myth of *in pluribus unum*, and for his version of the doctrine of *e pluribus unum*.)

We now have enough expository background to turn to criticism.

4.2.1 The nominalistic extensional theory of meaning.

To begin with, the extensional theory — in its simplest, nominalistic form — is the view that the meaning of a term is the object, or set of objects, the term refers to. Quine says that this theory fails since some meaningful terms have no determinate extensions, and some pairs of non-synonymous terms have identical extensions. There is no doubt that the extensional theory is false; in fact, it is *necessarily* false: false not only because, as a matter of fact, there are meaningful terms without determinate extensions, and pairs of non-synonymous terms with identical extensions, but because its falsehood is determinable on conceptual grounds alone and independently of non-semantic matters of fact. For extensions — that is, particulars or sets of particulars referred to by symbols — cannot be the meanings of symbols. Meanings are (semantic) properties of *symbols*; but extensions are not properties of symbols: they are what symbols refer to, and could exist even if there were no symbols. For example, Kaspar Hauser was not a semantic property of his name, and existed *before* he had any name whatever; but his name, one way or another, refers to him, and is meaningful though he no longer exists.

It may be, as I think is indeed the case, that the meaning of a term consists in the term's *denoting* a certain *nominal* property; but it cannot consist in any particular object, or set of objects, which partake of the property, or for that matter in the property itself. The extensional theory is therefore *necessarily* false; it is not a genuine account of meaning. So far, however, we have no reason to reject term-based semantics, for the classical denotational theory may be true; and it is worth noting that no classical term-based theorist ever held the extensional theory of meaning: only critics of term-based semantics have attributed it to them.

4.2.2 The realistic extensional theory of meaning.

Consider Quine's argument against the emended, realistic version of the extensional theory. This version is more interesting, for it relies on the existence of universals: it says that the meaning of a term, say, "Pegasus", is the property <Pegasus>; the meaning of "creature with a heart" is the property <creature with a heart>; *etc*. Quine's argument against this view comes roughly in three parts: *(i)* concerning the conceptual-scheme relativity of ontology; *(ii)* concerning the explanatory simplicity and power of a conceptual scheme; *(iii)* concerning the actual use of language and actual practices of positing universals. The point of my case against Quine will not be, of course, to defend the emended extensional theory, but to show that his arguments against it do not give any support to the holistic-*cum*-behavioural notion of meaning he offers us not just as an alternative to it, but as the natural outcome of its failure.

4.2.2.1 The conceptual-scheme relativity of what there is.

The positing of universals is bad metaphysics or ontology, Quine argues, because what entities (properties, particulars) there are depends on what *conceptual scheme* one uses to individuate entities. If one uses a *realistic* conceptual scheme, one recognises the existence of universals; but if one deploys a *nominalistic* scheme, then not; and since one cannot tell which of these schemes is true of the world, since it is impossible to decide absolutely and objectively what entities there are, and whether or not there are properties, the positing of absolute and objective universals is bad metaphysics.

The only conceptual schemes Quine takes into account are either a certain crude version of metaphysical *realism*, according to which there is an absolute and objective universal for each meaning-bearing term, or a certain crude version of metaphysical *nominalism*, according to which there are no universals for any meaning-bearing terms. But the alternatives need not be so stark. Notably, Quine's own brand of nominalism is not so extreme (see Quine (1969)): although he allows of no 'Platonic universals' existing independently of particular things, he does allow of there being *natural kinds*, or common properties of existent particulars. Also, Quine regards *sets*, like properties, as universals (see Quine (1953b: 114–15; 1960: 239)), and he is happy with there being sets. So, at least for natural-kind terms and set terms, his argument — by his own standards — fails to rule out the realistic extensional theory.

But even in regard of terms which purport to represent universals other than natural kinds or (if you think they are universals) sets, the alternatives Quine considers are too crude. This is because the only species of universal he takes into account are absolute and objective, *mind-independent* universals; and since these are empirically unaccountable, Quine concludes that the emended extensional theory cannot be correct. This again need not be so. Whatever one thinks of universals existing independently of particular things and of human minds, and whatever one thinks of natural kinds or even such abstract particulars as sets, there is still a plausible alternative which satisfies both the realistic posit of universals and the nominalistic tenet that *all that exists is particular*: namely, that there are *nominal* universals. Nominalism in fact derives its name not so much from the assertion that there are no universals, but rather from the view that universals, if there are any, are nominal and mind-dependent, having no existence apart from what is fixed by *tokens* of mental symbols. Such a view could be quite acceptable to Quine, had it not been for his desire for *mindlessness*; which desire is, perhaps, best attributed to that peculiar philosophical and scientific culture under whose influence he formed his opinions; the culture which attempted to screen off the mind from what there is, and to convince us that what appear to be thoughts, ideas, emotions, reasonings, *etc.*, are but ways of verbal and other behaviour (so that even the appearance of them is not an appearance, but a manner of moving and making noises). However this bizarre phenomenon came about, it prevented Quine and most other philosophers and scientists of that time from understanding that *term-based semantics*, the target of their polemical zeal, always held — in its classical form due to Locke and Descartes, among others — that the meaning of a term consists in its representing a *nominal universal*, not *real* or mind-independent universal (with some significant exceptions which were a matter of profound controversy). (See Book III of Locke's *Essay* for the best exposition of classical semantics.)

The lesson for us is two-fold. It may be that what entities can be *said* to exist depends on the conceptual scheme we use; but that conceptual scheme is fundamentally psychological rather than public; which in turn allows us to *speak of* mind-dependent universals, and so (to put it in my nomenclature) denote nominal properties. The emended extensional theory is *certainly* false; but Quine's argument from the relativity of conceptual schemes does not show it so, for the universals corresponding to terms could be nominal, which he overlooks. The other lesson is that the failure of the extensional theory of meaning, in any form whatever, gives us no reason to abandon term-based semantics and resort to Quine's behavioural holism; but more on term-based semantics in what follows.

4.2.2.2 The explanatory power and simplicity of a conceptual scheme.

Quine holds that even though there is no absolute and objective way to decide whether realism or nominalism is true of the world, it is possible to adjudicate between the two schemes on the grounds of such *pragmatic* criteria as simplicity and explanatory power; and that not only is nominalism simpler, it gains ontological simplicity without any loss of explanatory power. But is this true? The simplicity of a theory is not measured merely by the presence or absence of certain properties; but my main concern is with the claim that Quinian nominalism loses no explanatory power. The trouble with it is that, according to Quine's excessive empiricism, the *explananda* of theories are nothing but *experiences*; and since one cannot have, in principle, any experiences of abstractions such as universals, it appears to follow that one can abandon all reference to universals without losing any explanatory power, yet still agree with the realist as to what experiences there are on which occasion (*e.g.*, which houses or roses are red, *etc.*).

Firstly, barring phenomenalist science, the purpose of theorising in natural science is not merely to account for our experiences, but rather to explain the nature and functioning of the physical environment which occasions in us the experiences. We take our experiences as *evidence* for the obtaining of certain states of affairs in the environment. The point is that, in general, the purpose of theorising is not to explain the evidence: that would be putting the explanatory cart before the evidential horse; rather, we use experiential and whatever other evidence we may get to confirm or disconfirm theories about what states of affairs obtain in, and govern the working of, the environment.

Secondly, the claim that nominalism (of the Quinian sort) loses no explanatory power in abandoning all talk of properties fails entirely in the area of semantics, both linguistic and psychological. This counter-claim I will defend in the rest of this chapter, and especially in Chapter 5 (on behavioural and holistic semantics). For the moment, suffice it to note that if, following Quine and the spirit of his age, one screens off the mind from what there is, and adopts his brand of materialistic-*cum*-nominalistic ontology, there seems indeed no alternative but to turn to overt verbal and non-verbal behaviour for an explanation of representation and meaning; yet this consequence one should regard as a warning that something is wrong with the ontology, not as a bullet one has to bite!

4.2.2.3 The actual use of language, and actual practices of positing universals.

The nominalistic conceptual scheme is, Quine argues, not only equal in explanatory power to the realistic scheme, but superior; and this because nominalism comports better with facts about the actual use of language and actual practices of positing and individuating properties. In short, according to Quine, nominalism offers a better (ironically, *the only correct*) account of linguistic meaning. The account is, to repeat, that since language is a social art which we acquire solely from 'intersubjectively available cues as to what to say and when', it follows that 'there is no justification for collating linguistic meanings, unless in terms of our dispositions to respond overtly to socially observable stimulations' (*op. cit.*). The nominalistic feature of this behaviour-based account of meaning consists in that the 'socially observable stimulations' and 'intersubjectively available cues' need and can be bound only by *resemblance*, not identity in any respect, so that semantic uniformity emerges *from the many*, rather than being implicit *in the many*: the democratic creed *"e pluribus unum"* triumphs over "*in pluribus unum*".

There are, among others, two reasons why Quine does not get what he wants. Firstly, his characteristic argument from language-learning to behavioural semantics is surely *invalid*. It may be true that language is a social art; it is no doubt so in large part. It may also be that 'in acquiring language we have to depend entirely on intersubjectively available cues'; I think language acquisition need not *entirely* depend on 'intersubjectively available cues', but there is no harm in supposing so for the sake of argument. From this, however, it does not follow that 'there is no justification for collating meanings unless in terms of dispositions to respond overtly to socially observable stimulations'. To suppose it does is to conflate the concerns of the theory of language-acquisition with the concerns of semantics. As far as semantics is concerned, it matters *that* one acquires language, and it matters *how* one's linguistic symbols manage to represent or mean; but it does not matter *how* one acquires language. As far as the theory of language-acquisition and, more generally, concept-acquisition is concerned, it matters *how* one acquires language and concepts, but one need not worry about how linguistic symbols manage to be meaningful. I do not suggest that semantics and the theory of language acquisition cannot bear onto one another's concerns, only that they are not one and the same; and I think much of the foundational work in semantics can be done independently of the theory of language-acquisition, and conversely. The root of Quine's fallacy lies in that he restricts the resources of semantical theory to the resources of behavioural psychology and linguistics; but there is no good reason why that restriction should be in place.

Secondly, the nominalistic facet of Quine's behavioural account of meaning gives way as soon as one abandons the argument from language-acquisition, and with it the conclusion that meanings can be collated only from verbal behaviours *vis-à-vis* stimulatory experiences. For Quine's *mind-less*, materialist nominalism requires a behavioural account of linguistic meaning; if behavioural holism turns out unjustified and untenable, as I will show in Chapter 5, it will follow that nominalism of the Quinian variety cannot match the explanatory power of conceptual schemes which accommodate some, perhaps several sorts of universal.

We shall next look at some of Quine's problems with the other alternatives of Plato's beard; *viz.*, the view that meanings are either mental or mind-independent semantic universals. Having surveyed these, we shall turn into Quine's home territory, the *mind-less* nether-world of semantic behaviourism and behavioural holism.

4.3 Mental and Mind-Independent Semantic Universals

Reflect again on the theory that the meaning of a term is the object, or set of objects, the term refers to; that is, the simplest version of Plato's beard. As we have already seen, this account is not acceptable because, amongst other reasons, some meaningful terms have no determinate extensions, and some pairs of non-synonymous terms have identical extensions. Consider now the following emendation one might put forth in order to reclaim Plato's beard, though not in the form of the extensional theory, despite these arguments. One might say that the meaning of a term is a *semantic universal* — an *idea* (or concept, sense, intension), either *mental* or *mind-independent* — which determines but is not identical with the term's extension. Thus, one might say that the meanings of "creature with a heart" and "creature with a kidney" are, respectively, the ideas — whether mental or mind-independent — **creature with a heart** and **creature with a kidney**; and that these ideas determine identical extensions for the terms. Similarly, one might say that the meaning of "Pegasus" is the idea **Pegasus**, and that **Pegasus** determines the extension {Pegasus} (which happens to be empty). Quine finds this emendation unsatisfactory, for the following reasons.

Firstly, Plato's beard so emended is bad metaphysics since it presupposes the existence of universals, specifically, semantic properties. The reasons why this is bad metaphysics are the same as those given in Section 4.2, and need not be repeated. (See Quine (1948: 11) for a sketch of Quine's nominalistic qualms about semantic universals.)

Secondly, the emendation is unsatisfactory because it begs the question of semantic individuation, in that the notion of an idea (concept, sense, intension) is no less in need of an explication than the notion of meaning;

in fact, it *is* the notion of meaning. Quine says: "The evil of the idea idea is that its use, like the appeal in Molière to a *virtus dormitiva*, engenders an illusion of having explained something. And the illusion is increased by the fact that things wind up in a vague enough state to insure a certain stability, or freedom from further progress" (Quine 1953a: 48).

Thirdly, the emendation is unsatisfactory because it presupposes that *complex* ideas comprise necessary and sufficient conditions for the membership in the extensions of their associated terms; or better, that terms signifying complex ideas are semantically definable from terms signifying simple ideas (whatever the simple ideas might be). However, terms allegedly signifying complex ideas are not definable from any basis of semantic simples; hence complex ideas, if there were any, could not comprise necessary and sufficient conditions for the membership in the extensions of their associated terms, and so could not determine the extensions of the terms. (See Quine (1969d; 1966a: 208; 1966b: 54–55) for some objections against conceptual definability.)

Quine takes it that these arguments show that Plato's beard — whether in the form of the theory that meanings are mental universals, or in the form of the theory that they are mind-independent universals, must be given up. Further, he takes the arguments to show that we must abandon the notion of the meaning of a term as that which determines the term's extension, and dissociate the meaningfulness of terms from their determination of extension: "When the cleavage between meaning and reference is properly heeded, the problems of what is loosely called semantics become separated into two provinces so fundamentally distinct as not to deserve a joint appellation at all. They may be called the *theory of meaning* and the *theory of reference*" (1953c: 130). Notice that the three arguments together establish the need, insofar as concerns Quine's position on Russell's problem, for the alternative, non-Russellian solution to the problem; for Russell's solution assumes the *definability* of terms, or denoting phrases, which according to Quine cannot be had. Quine's solution is therefore bound to rest on the rejection of the notion of meaning as extension-determiner. (Compare the foregoing arguments with those of Wittgenstein (1953). For example, see (1953: §§ 20, 32–36, 673–693) for his denial of semantic mentalism, and his adoption of behavioural-*cum*-holistic semantics; also, see (1953: §§ 39–44) for a discussion of Russell's problem, and for his abandonment of the traditional notion of meaning as extension-determiner; and see (1953: §§ 46–50) for his rejection of Plato's beard on the grounds that, inasmuch as it presupposes the existence of absolute and objective, conceptual-scheme independent properties, Plato's beard is bad metaphysics.) Let us now examine Quine's arguments.

4.3.1 The bad metaphysics charge.

Like the emended extensional theory, the mentalistic or platonistic account presupposes the existence of universals — namely, semantic properties —

and is therefore bad metaphysics; so Quine argues. I would agree that positing semantic universals is bad metaphysics only if one were to think of them as properties not inhering in any particular. But this, of course, is not necessary. One can, and should, think of semantic properties as properties of *symbols*, consisting in the relations of denotation between the symbols and certain nominal properties (*i.e.*, nominal *non-semantic* properties; the *denotata* are not to be conflated with the relations of denotation). One could object that this leaves us still with non-inherent universals: symbols are types of entity, and even the nominal properties are types which need not be instantiated by any particular. However, there is no difficulty in sorting this out.

According to the Classical Theory of Mind, the primary bearers of meaning are *ideas*, or *tokens* of mental symbols: each idea being, to put in a Lockian way, a *particular existence*. Although Locke often employs the term "*Idea*" to stand for either types or tokens of symbols, this is merely for convenience; at several places in the *Essay*, he makes it clear that each idea is a particular occurrence in the mind, and in general that everything that exists is particular. To put it emphatically, an idea is a *token* of a symbol, a mental occurrence of a certain type. Accordingly, semantic properties of ideas are common aspects of particular existences; and, *in that regard*, they do not differ in ontological status from *natural kinds* (which Quine has no qualms about). Semantic universals are thus *inherent* aspects of particulars, not free-floating universal entities in a mysterious realm of properties without a bearer. For instance, the notion (that is, the semantic universal) BLUE is a property of the idea **blue**, consisting in that **blue** denotes the nominal property [blue]; and **blue** is a token of a mental symbol type.

Further, the *denotata* of ideas — *i.e.*, the nominal (non-semantic) properties denoted by ideas — need not be instantiated by any particular object in the environment; *e.g.*, nothing partakes of the nominal property [the present King of France]. So it would appear that nominal properties need not inhere in any particular. But again, not so. The nominality of the property denoted by an idea consists in that the identity of the property is fixed not by the environment or noumenal world, but by the idea itself; more generally, by the mind. The identity of [blue] is fixed by the idea **blue**; the identity of [the present King of France] is fixed by the constituent simple ideas of the complex idea **the present King of France**; *etc.* One might wonder how an idea determines the identity of its *denotatum*. This is not a problem in the case of a complex idea such as **the present King of France**; the identity of [the present King of France] is fixed by *description*, from the semantically simple constituents of the idea. By contrast, in the case of a simple idea such as **blue**, we must say that the identity of [blue] is determined by the *form of consciousness* which is essentially and inseparably associated with each tokening of the idea (that is, by the *sensation* of blue one has in thinking the idea). More generally, the idea **blue**, which is a particular existence in the mind, has several sorts of properties: it has a *semantic property* consisting in its denoting the nominal essence [blue]; it has a certain kind of *syntax* which allows it to enter causally into the formation of complex ideas and into mental processes; and it has a *form of consciousness*, which is the sensation one experiences in thinking the idea. What is most important for our present concerns, the identity of the nominal essence [blue] — or the complex nominal essence [the present King of France] — is always determined by a *particular existence*: namely, by the simple or complex idea which denotes it and determines its identity, whether or not any particular object in the environment instantiates the property. In brief, we do not commit ourselves to non-inherent universals in speaking of such nominal properties as [the present King of France], *etc.*

Clearly, the ontology of semantic universals and nominal essences sketched above is *nominalistic*: it holds that all that exists is particular. Yet it is a *mentalistic nominalism*, not the *mind-less* nominalism of Quine. What prevented Quine from sorting out such ontological problems for himself was precisely the research policy of screening the mind off from what there is, a policy which was not of his own making, and which he accepted as uncritically as before him Frege accepted imagistic psychology and after him Fodor accepted computational psychology. To return to Quine's objection that positing semantic universals makes for a bad metaphysics: quite on the contrary, classical mentalistic nominalism does not have a difficulty with semantic universals, though Quine's mind-less nominalism surely does.

4.3.2 The *virtus dormitiva* charge.

Quine argues that the mentalistic or platonistic account of meaning is like a *virtus dormitiva*, in that it begs the question of semantic individuation, since the notion of an idea is no less in need of an explanation than the notion of meaning. I agree that, in Quine's formulation, the account is circular; but Quine is battling a straw opponent. The Classical Theory of Mind says, at least, that the meaning of a *public* symbol derives from the meaning of an associated token of a *mental* symbol, or idea; and that the meaning of the idea consists in its denoting a nominal universal, the identity of which is determined by the idea itself. So the meaning of "yellow" consists in the term's denoting the nominal essence [yellow]; and the identity of [yellow] is fixed by the idea **yellow**. As for complex ideas, their meaning likewise consists in denoting nominal essences which are fixed by the semantically simple constituents of the ideas. I will spell out this account in Chapters 7 and 9; but it is already clear that the classical posit of ideas as meaning-bearing entities is not a *virtus dormitiva*; ideas, as tokens of mental symbols, are given a denotational account of meaning, much as tokens of public symbols are. The fact that we do not know the answers to some of the great questions of CTM, concerning the nature of ideas and of consciousness, certainly does not invalidate the classical posit.

4.3.3 The undefinability charge.

Quine argues that the mentalistic or platonistic account of meaning is unacceptable since it presupposes that terms signifying complex ideas are definable, by necessary and sufficient conditions, from terms signifying simple ideas (whatever the simple ideas might be); but terms allegedly signifying complex ideas are not so definable from any basis of semantic simples, hence mentalism or platonism cannot be correct. This argument raises an important issue about the size and character of the basis of semantically simple mental symbols, and about the mind's composition of complex symbols.

We should note, to begin with, that the plausibility of the argument depends on two factors: firstly, whether the basis of semantically simple ideas is rich enough to be adequate for defining all complex notions we can express; for if it is too small or too tendentious, the argument is very plausible; and secondly, whether the mind's composition of complex ideas must be rigorous enough to ensure that each complex idea comprise necessary and sufficient conditions for the membership in the extension of the associated term; for if it need not be so rigorous, the idea need not determine the extension.

For Quine, the argument is conclusive since the only basis of semantic simples he takes seriously, following verificationist attempts at definitions, is *sensory*, comprising only terms standing for simple sensory properties, supplemented by set-theoretic and quantification terms; and since hardly any terms are definable from the sensory-*cum*-set-theoretic basis, since that basis is not rich enough to be adequate for defining most complex notions we can express, he rules out definability altogether, and with it term-based semantics. Further, Quine takes it that term-based mentalism or platonism, if it were correct, would rely on the Conservative rule of semantic individuation, that a difference in extension implies a difference in meaning. In the case of complex ideas, this would require that each complex idea comprise necessary and sufficient conditions for the membership in the extension of the associated term. But this is hardly ever the case: most notions people can express with their public terms fall far short of comprising such necessary and sufficient conditions, which shows that what the terms mean does not determine the corresponding extensions. Therefore term-based mentalism (or platonism) cannot be correct.

However, neither of Quine's suppositions holds: the basis of semantically simple ideas need not be and, it is certain, is not sensory-*cum*-set-theoretic; nor does mentalism rely on the wrong principle of extension-meaning supervenience, or require that each complex idea comprise necessary and sufficient conditions that determine the corresponding extension.

Firstly, Locke's empiricist basis of simple ideas is larger than that of his verificationist descendants; it includes not only *ideas of sensation* but also *ideas of reflection*, or mental signs representing cognitive and emotive operations such as believing, desiring, *etc.* In addition, Locke regarded some ideas as of *more than one sense*, other as of *both sensation and reflection*; for example, the ideas of unity, existence, extension, duration, and so on. With Locke's basis, one has a much better prospect for composing all complex notions we can express.

Secondly, one of the characteristic features of the Classical Theory of Mind is that it does not accept the wrong principle of semantic individuation that a difference of extension requires a difference of meaning; in fact, it *separates* (much as Quine himself does) reference, or determination of extension, from meaning, or denotation of nominal properties. To put it in a more classical style, CTM regards meaning as the representation of a *nominal world*, and dissociates this from a representation of the *noumenal world*; and it regards the latter as having to do with *knowledge*, not *meaning*. So classical semantic mentalism does not require that each complex idea be uniformly and strictly definable to comprise necessary and sufficient conditions that fix the extension of the associated term; in fact, on some versions of CTM, no ideas, simple or complex, can ever determine their extensions (witness Kant's version). More amiable versions, such as Locke's, do allow that this happens to some extent with simple ideas, and that it *can* happen with complex ideas, though not without a great deal of cognitive effort by the mind, to compose its complex ideas so as to represent the world veridically.

But, one might ask, if extension-meaning supervenience is not a criterion of semantic individuation for classical mentalism, so that *complex ideas* need not comprise necessary and sufficient conditions for the membership in their extensions, how does CTM fix the identity of meaning for words expressing complex ideas? As for *simple ideas*, these are the same and invariable for the human mind. (This is not to say that each individual mind must have the same range of simple ideas; the range of ideas it *can* have is the same, but what simple ideas it actually acquires depends on what experiences it has had; an individual blind from birth will not have simple ideas for which visual experiences are necessary, but individuals with similar experiences will have a similar range of simple ideas.) Thus, when two individuals use a word to express the same *simple idea*, their tokens are synonymous. As for complex ideas, though, these may and often do differ from person to person, and — for each person — from time to time. Here the classical rule is that meaning is *individualistic*: one means with a token of a word what one *has in mind* as an individual speaker; that is, the idea one uses the word to express or stand for *on that occasion*. So, if Jane's idea of gold is **yellow precious metal**, and John's idea is **yellow malleable metal**, then Jane's meaning of "gold" consists in denoting the nominal property [yellow precious metal], whereas John's meaning consists in denoting [yellow malleable metal]. To make communication possible and successful, individual minds must find a way of using public words, to put it in a Lockian manner, in *common acceptation*; and, as a matter of fact, they do more or less succeed. There are specialists on gold in most communities; their complex ideas of gold are more involved than those of laity, and, for them, the nominal essence [gold] perhaps more closely approximates to the *natural kind* <gold>. Perhaps even, for some experts, [gold] could very nearly match up, or correspond, to <gold>. But this does not make experts any better than laity in respect of *meaning*; their ability to mean is the same as laity's, though their *knowledge* be ever so perfecter.

The consequences of uncritically accepting the wrong principle that meaning fixes extension are dire. If one is a mentalist like Fodor, and, like him, one accepts that principle, one is driven by arguments such as Quine's to hold that all *prima-facie* lexical mental symbols — that is, symbols corresponding to the vocabulary of a public language such as English — are semantically simple, unlearned and innate; and this in the hope that, by avoiding the charge that complex ideas must comprise necessary and sufficient conditions for the membership in their extensions in order to work as extension-determiners, one will be able to give a *referential* account of meaning, and a *nomic* account of reference, for the simple symbols. Else, like Fodor (1994), one is driven to accept that most lexical mental symbols are complex, but still to run the referential account of meaning and nomic account of reference for the complex symbols, as though they were unanalysable units. These are desperate measures of a mad rationalism; and there is only one cure for the disease, to embrace the Classical Theory of Mind, and reject once and for all that old sorcerer's notion that meaning fixes extension, and that the natural world itself contributes to the identity of meaning, so that if you mean very ardently, if only you tap into the right meaning beyond appearances, you will thereby know how the world is, and make yourself unto God, ruling the world as you please by meaning as you please. True, there is another remedy for the *idée fixe*, prescribed by our behaviourist *medicinæ universæ doctor*, which is to go *mind-less*; but that really amounts to chopping off the patient's head.

Chapter 5

Radical Empiricism II

5.1 Two Major Routes to Semantic Holism

We have so far discussed, in relation to semantic holism, what Quine regards as the minor reductionist alternatives: the extensional theory of meaning, and the theory that meanings are either mental or mind-independent semantic universals. In this chapter, we shall turn to the major avenues by which Quine arrived at holism: semantic verificationism and behaviourism. Quine argues that both verificationism and behaviourism fail, and that the way they fail leaves open only behavioural holism as an account of meaning. Very generally, verificationism fails because it rests on the assumption that meanings are attributed to (not terms but) *sentences* one-by-one, in isolation from whatever larger linguistic and theoretical contexts the sentences occur in; behaviourism fails because, like other variations of semantic reductionism, it rests on what Quine takes to be the realistic myth of *in pluribus unum*, the view that, for every meaningful term, the entities the term applies or refers to must have some property in common. Let us now examine these arguments in detail.

5.2 The Route from Verificationism to Holism

We have seen in Chapter 3 that, according to semantic verificationism, the primary bearers of semantic properties are *sentences* — or better, *statements*, as tokens of sentences — taken one-by-one, in isolation from whatever larger linguistic contexts they occur in. Quine argues, to the contrary, that the primary bearers of meaning must be the larger linguistic contexts themselves:

> The idea of defining a symbol in use was, as remarked, an advance over the impossible term-by-term empiricism of Locke and Hume. The statement, rather than the term, came with Frege to be recognized as the unit accountable to an empiricist critique. But what I am now urging is that even in taking the statement as unit we have drawn our grid too finely. The unit of empirical significance is the whole of science. (Quine 1951: 42)

His argument runs as follows. If statements were bearers of orthodox extension-determining and truth-condition-determining meanings one-by-one, in isolation from their contexts, then it should be possible to *confirm* (verify, test, epistemically evaluate) the statements one-by-one; that is, in isolation from their contexts and with respect to their determinate truth-conditions. *A fortiori*, if statements were bearers of meaning one-by-one, and if the meaning of a statement were the method of its confirmation, then statements should be confirmable one-by-one. But, Quine argues, *no* statements are confirmable one-by-one: "... our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (1951: 41). The reason is that confirmation is *holistic*: one cannot evaluate a statement either *solely* from its (putative) correspondence to its determinate truth-condition, or *solely* from its meaning; one must draw supporting evidence for the evaluation of the statement (not only from its putative truth-condition or from its meaning but, in addition) from the truth or falsity of other statements and theories in science, *and* these other statements and theories may belong to any, however remote, branch of science. Why is this so?

> The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Reëvaluation of some statements entails reëvaluation of others, because of their logical interconnections — the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having reëvaluated one statement we must reëvaluate some others, which may be statements logically connected with the first or may be the statements of logical connections themselves. But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reëvaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.
>
> If this view is right, it is misleading to speak of the empirical content of an individual statement — especially if it is a statement at all remote from the experiential periphery of the field. (Quine 1951: 42–43)

Thus no statement is epistemically evaluable either solely from its correspondence to empirical facts or solely from its meaning; the evidence for the evaluation of a statement must be drawn from other statements and theories, and these statements and theories may belong to any branch of science. In short, epistemic evaluation (confirmation, verification) is holistic.

It follows, according to Quine, that statements cannot be bearers of truth-condition determining semantic properties one-by-one, in isolation from whatever larger linguistic contexts they occur in; and since the argument applies to, and is iterable for, any isolable linguistic form, however large, it follows that no linguistic forms are bearers of the orthodox extension and truth-condition determining semantic properties; that is, what follows is *semantic nihilism* with respect to the Conservative notion of meaning as extension-determiner; and it follows that the fundamental 'unit of empirical significance' can be only the whole of science (or a large portion of it).

Quine's conclusion that meaning is holistic rests on the premiss of confirmation holism, the doctrine that one *must* draw supporting evidence for the confirmation of *any* statement from the truth or falsehood of other statements, *and* these other statements may belong to *any* branch of science. Quine's premiss is thus identical to Carnap's conclusion that confirmation is always under-determined by evidence, a matter of pragmatic decision (see Section 3.4). Quine's own argument for confirmation holism is, however, different from Carnap's; it is the argument from the alleged epistemic reëvaluability or *revisability* of statements, summarised in the foregoing quotation. I will turn to the argument from revisability anon. But firstly let us note the connection between Quine's and Carnap's positions on meaning and confirmation: in brief, Quine agrees with and entirely adopts Carnap's conclusion that, for all *synthetic* (*i.e.*, contingent) statements, *confirmation* is under-determined by empirical evidence, a matter of pragmatic decision rather than a determinate method; but Quine proceeds to draw the further conclusion, already implicit in Carnap's position, that *meaning* is likewise indeterminate and thus holistic; in addition, Quine generalises Carnap's position to include *analytic* as well as contingent statements: *i.e.*, he insists that even putatively analytic statements are confirmable at best holistically and therefore have no *determinate* semantic properties, or meanings.

We shall next focus on Quine's argument from revisability, which says that since all statements are revisable or reëvaluable depending on revisions elsewhere in science, it follows that confirmation and consequently meaning are holistic; and, in view of Quine's developments on Carnap's position, we shall divide the discussion into two sub-sections, so as to treat of synthetic (contingent) statements and of analytic statements under distinct headings.

5.2.1 Synthetic statements.

To be clear about the target argument, here it is again. If statements were bearers of orthodox extension-determining semantic properties one-by-one — *a fortiori*, if the meanings of statements were the methods of confirmation of the statements — then they would have to be confirmable one-by-one, with respect to their determinate extensions and truth-conditions. But no statements are so confirmable, since confirmation is holistic: one must draw evidence for the confirmation of any statement from the truth or falsity of other statements and theories, and these statements and theories may come from anywhere in science. Hence no statements are bearers of orthodox extension-determining semantic properties; meanings, if they are anything, must be holistic. The reason why confirmation, and so meaning, is holistic is that any statement is epistemically revisable depending on revisions elsewhere; no kind of evidence is ever *sufficient* to prove any statement conclusively.

Regarding *synthetic* (contingent) statements, my objections to this argument will come in three stages: *(i)* Quine relies (when it suits his verificationist leaning) on a phenomenalistic conception of natural science; but clearly phenomenalism need not and should not be accepted; *(ii)* the thesis of universal epistemic revisability does not hold; *(iii)* Quine's inference from confirmation holism to semantic holism likewise does not hold.

5.2.1.1 Phenomenalism in natural science.

We noted in Section 4.2 that, for Quine, the *explananda* of scientific theories are nothing but *sensory experiences* (however badly this comports with his overall physicalism and behavioural psychology). In the foregoing quotation, which supposedly shows his revisability thesis, Quine compares science to 'a field of force whose boundary conditions are experience'; he goes on to say that 'a conflict with experience at the periphery occasions readjustments in the interior of the field', and that 'the total field is underdetermined by its boundary conditions', so that consequently 'it is misleading to speak of the empirical content of an individual statement'. Quine's position is, more explicitly, that if statements of natural science were bearers of the Conservative truth-condition determining meanings, the truth-conditions would have to be *sensory* states of affairs. His conclusion, that no statements are confirmable solely from their correspondence to empirical facts, so that confirmation is holistic, then clearly follows; for no statements — taken as public tokens of the same syntactic type, made by different persons or at different times — have determinate and universally fixed *experiential truth-conditions*. But of course this view of natural science is untenable. More plausibly, the truth-values of statements in natural science and common-sense observation depend on what particular facts actually obtain in the environment, and on the nature and functioning of the environment which occasions in us our sensory experiences. We take our experiences as *evidence for* those facts, not as a phenomenal reality fixing the truth-values of our statements.

5.2.1.2 Universal epistemic revisability.

The problem we are now facing is this: are there any synthetic statements which are confirmable solely from their correspondence to empirical facts, and independently of any other statements? Some terminological clarification will be helpful before we proceed. Synthetic statements include all contingent statements; and of these, we may concede, *most* are confirmable only holistically. So syntheticity cannot be identified with confirmability solely from correspondence to empirical facts and regardless of other statements. Further, some synthetic statements might be *a priori* and hence *not contingent*. We shall treat of *a priori* synthetic statements (or propositions) in Chapter 9; at present, we should note only that such a statement would be, for evaluative aims, void of empirical parts, and confirmable solely from its non-empirical, *a priori* constituents and regardless of any other statements. There are, therefore, two ways a synthetic statement might be confirmable *non-holistically*: it might be an *a priori synthetic* statement, in which case it would be confirmable with *a priori* certainty from its constituent parts; or it might be a *contingent synthetic* statement, which is yet *confirmable solely from its correspondence to an empirical fact*, by taking certain sensory experiences as sufficient evidence for the fact. We can still allow that most contingent synthetic statements are confirmable only holistically; but the problem we are about to consider in this sub-section does not concern these; similarly, it does not concern *a priori* synthetic statements. Our problem is only with synthetic statements which might be confirmable solely from their correspondence to empirical facts, by taking certain experiences as sufficient evidence for the facts, and regardless of any other statements.

As we have seen, Quine obfuscates this issue by tacitly assuming that empirical facts must be sensory, experiential, phenomenal states of affairs; whereas sensory experiences had better be regarded as *some* evidence for the facts, not as the facts themselves. This granted, the problem we wish to answer is: is experiential evidence ever *sufficient* for the evaluation of a statement? It is plausible to interpret Quine's revisability thesis (*i.e.*, the claim that any statement is reëvaluable or revisable depending on evaluations of other statements and theories in science) as implying that experiential evidence is never sufficient for the evaluation of any statement, and hence that confirmation is holistic. This is the gist of Quine's argument.

Notice first that the only synthetic statements Quine seriously considers are *theoretical* statements of natural science, whether they be 'the most casual matter of geography and history' or 'the profoundest laws of atomic physics'. He rests the argument entirely on his account of natural science, and then generalises, *ex cathedra*, without further argument, to cover all statements. Everyone agrees that theoretical statements of natural science are confirmable only holistically, and that experiential evidence alone is never sufficient to confirm such statements. This also accords with classical epistemology: for example, Locke's position is that most propositions, those of natural or experimental science in particular, are demonstrable "only as they more or less agree to Truths that are established in our Minds, and as they hold proportion to other parts of our Knowledge and Observation" (1975: IV, XVI, 12).

But are there non-theoretical statements, common-sense observational statements about perceptible objects, that are confirmable solely with respect to empirical facts, by taking certain experiences as evidence for the facts? To be sure, sole experiences are often not sufficient for the evaluation of common-sense observational statements: the viewing conditions might be bad; one might need to use more than one sense without being in a position to do so; *etc*. But the issue is whether such statements are confirmable solely from their correspondence to empirical facts *under favourable observational circumstances*. There is a strong *prima-facie* case for the affirmative, since we all daily rely on such statements. This alone should put the onus of proving the contrary on the holists; alas, this is hardly to be expected, given the *status quo* of Academic holism. Here is an argument which might perhaps change someone's mind.

The following requirement on epistemic reëvaluation should be acceptable to all: that when a statement, *qua* token of a type of sentence, is reëvaluated as a result of evaluations of other statements, the reëvaluated statement must be a token of the same type as the original; that is, it must be the same type of statement. Otherwise no reëvaluation has occurred, only a replacement of one statement with a statement of a different type. Specifically, this requires that the reëvaluated statement must be of the same *semantic* type as the original: they must have the same meaning. It would not do that the reëvaluated statement be only syntactically identical to the original, but different in meaning. For example, the statement "I see a bank", which might happen to be subject to reëvaluation, must remain semantically invariant under a legitimate reëvaluation; else we might begin with a statement about a financial institution, and finish with a statement about a riverside. In short, our requirement for a legitimate reëvaluation of a statement is that, though the reëvaluated statement be assigned a different truth-value than the original, it must be the same type of statement, meaning and all.

We may suppose without being tendentious that theoretical statements of science can and do satisfy this requirement. They would not satisfy it if semantic verificationism were true; for then a change in the assignment of truth-value would imply a change in proof; and a change in proof would imply a change in meaning. But, barring as we do verificationism, theoretical statements can and do stay semantically invariant under reëvaluation. In contrast, some common-sense observational statements cannot satisfy the requirement. Consider, for instance, a statement such as "this is yellow", asserted on a particular occasion of a particular yellow marigold. Suppose one evaluates the statement as true by taking one's experiences of yellow, under favourable observational circumstances, as evidence that the thing is yellow. Lastly, suppose one subsequently tries to reëvaluate that statement because of reëvaluations of some other statements or theories. What one *means* by "yellow" is, I take it, not the real constitution of the marigold that makes it yellow. That is, one does not mean the natural kind <yellow-of-marigold>; one may be said to *refer* to <yellow-of-marigold>, but what one *means* or *denotes* is a nominal kind [yellow], the identity of which is determined by one's experience, or form of consciousness, ⟦yellow⟧. Accordingly, in evaluating the statement "this is yellow" as true, one asserts nothing about the natural kind <yellow-of-marigold>; what one asserts is that the thing partakes of the nominal kind [yellow]. Now, in purporting to reëvaluate the statement because of reëvaluations of some other statements or theories, one presumably implies that the state of affairs the statement represents, namely, [this is yellow], was as a matter of fact not instantiated on the occasion, and that some other state of affairs in fact obtained. But what could this other state of affairs, that was in fact instantiated, be? If you take it to be some *real essence* of the marigold, or generally any real aspect of the environment, you will succeed in keeping the meaning of "this is yellow" invariant under the reëvaluation, but the 'reëvaluation' will not concern that statement; it will concern some other statement or theory such as a scientist, if anyone at all, might make. If, on the other hand, you take it to be some *nominal essence*, different from [this is yellow], you will not keep the meaning of the original statement constant: the reëvaluated statement will be semantically different from the original, though syntactically identical. Either way, no reëvaluation of the statement "this is yellow" has occurred.

Hence it is clear that some — not all, but quite a few — synthetic statements we make are evaluable solely from their correspondence to empirical facts, by taking certain experiences as sufficient evidence for the facts, and regardless of evaluations or reëvaluations of other statements or theories. I stress that the issue is not whether one can rule out every possibility of error in confirming common-sense observational statements; rather, it is whether such statements, once confirmed under good circumstances, are subsequently revisable as a result of revisions of sundry statements or theories elsewhere. I have argued to the contrary, on the grounds that, in purporting to reëvaluate such statements, one either reëvaluates some other, theoretical statements concerning the nature of the environment, or else one 'reëvaluates' the observational statements only at the cost of changing their meaning. So far, then, Quine's argument from revisability fails to show that no statements can be confirmable one-by-one, in isolation from their contexts, and therefore fails to establish confirmation holism.

5.2.1.3 The inference from confirmation holism to semantic holism.

I will now show that, as regards synthetic (contingent) statements, semantic holism would not follow even if confirmation holism were true. If Carnap's semantic verificationism were true, then the inference from confirmation holism to semantic holism would hold: if meanings were ways of verification, and verification were always holistic, then meaning would also be holistic. This is most likely how Quine initially arrived at semantic holism. But verificationism is false; hence one could not get to semantic holism *via* verificationism even if confirmation holism were true.

Perhaps, however, on a more charitable reading of Quine, one might get to semantic holism from confirmation holism without subscribing to verificationism. One might argue, as I said in the beginning of Section 5.2, that if meanings were the extension-determiners and truth-condition-determiners of statements, then statements should be evaluable one-by-one, with respect to their determinate truth-conditions and independently of their contexts; but if no statements were so evaluable (*i.e.*, if confirmation holism were true), then no statements could be bearers of the Conservative extension and truth-condition determining semantic properties, and meaning would have to be holistic.

So interpreted, is the inference from confirmation holism to semantic holism correct? It would be, if classical semantic properties were required to be extension and truth-condition determiners. But the Classical Theory of Mind never held the wrong principle that meanings are extension and truth-condition determiners; in fact, that rule became a part of Conservative thought only with the advent of Analytic Philosophy, and has since caused no end of confusion. Hence the inference from confirmation holism to semantic holism does not go through, even on the charitable reading, provided we reject the rule of extension-meaning supervenience.

There is another way to see why the inference from confirmation holism to semantic holism fails (as regards contingent statements). Suppose confirmation holism is true; that is, that experiential evidence is never sufficient for the evaluation of any statement, and that one must draw supplementary evidence from evaluations of other statements and theories, which may come from any branch of science. Still, semantic holism does not follow. The reason is that holism need not and does not hold even of statements confirmable, as a matter of fact, only holistically. The statement "stars are hot gaseous bodies each of which is at least four light-years away from the sun", for instance, is not evaluable solely from its correspondence to the fact that stars are hot gaseous bodies each of which is at least four light-years away from the sun, by taking certain experiences as evidence for that fact; rather, its confirmation is holistic: one must seek evidence for its confirmation from other statements and theories, which may come from sundry unexpected quarters in science. Yet it does not follow that the meaning of the statement is holistic; its meaning may and does consist in its representation of a nominal state of affairs, *viz.*, [stars are hot gaseous bodies each of which is at least four light-years away from the sun]; and that nominal state of affairs is fixed by a proposition, or token of a sentential mental symbol, in the mind of the speaker who uses the statement.

5.2.2 Analytic statements.

Epistemic holism for contingent statements, as we have seen, does not imply semantic holism. But it does rule out there being contingent statements evaluable only from their correspondence to empirical facts, by taking certain experiences as evidence for the facts, and regardless of any other statements. Further, epistemic holism *in general* — *i.e.*, for all kinds of statement — rules out there being any *analytic* statements, or statements evaluable *solely* from their own meaning and independently of any other statements; for if no statements were evaluable unless by drawing evidence from other statements, surely there could not be any statements evaluable solely from their own meaning. But then, supposing, for the sake of argument, that general epistemic holism is true, and accepting that it rules out semantic analyticity, it follows that one can infer semantic holism from epistemic holism after all. This is because analyticity requires that terms and statements should have determinate (non-holistic) semantic properties; conversely, if terms and statements had determinate semantic properties, then — granted that language is generative, so that one cannot simply cut all analyticities out of a language whilst keeping the rest of the language intact — *some* statements would turn out evaluable solely from their own meaning and so analytic. Hence we have it that if no statements were analytic, as would hold if epistemic holism were true, then any reductionist notion of meaning would have to be incorrect, and semantic holism would have to follow. So there is a way of getting from epistemic holism to semantic holism, provided epistemic holism is general, ruling out analytic statements.

Not surprisingly then, Quine argues that there are no analytic statements. His argument here is the same argument from epistemic revisability we have already dealt with; it will be helpful, though, to review it with special attention to the problem of analyticity:

> [Assuming revisability] it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may. Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the [experiential] periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?
> (Quine 1951: 43)

Note again that Quine's notion of a synthetic statement is a truncated one inherited from the verificationists; and his various notions of an analytic statement (see (1951: *passim*)) are not any better. It is no wonder he should have trouble with sorting out what analytic and synthetic confirmation involves. But let us turn to his argument: it says that since all statements are revisable as a result of theory change, there can be no statements which 'hold come what may', or analytic statements.

What Quine needs in order to show that there are no analytic statements is to show that every purportedly analytic statement is reëvaluable as a result of reëvaluations elsewhere in science, *whilst its meaning is held constant and invariant*; otherwise no genuine revision has occurred, only an exchange of one statement for another. But that is what he does not and cannot get. For example, take the statement — or statemental function — "α or not α". Quine says the statement is subject to revision because of theory change in quantum physics; but all he shows is that, in some domains of discourse, the syntactic forms "or" and "not" are given different meanings. This 'revision' rests on a change of meaning; semantic revision precedes the alleged epistemic revision.

Further, contrary to Quine's claim that there is no difference in principle between revising a logical truth such as "α or not α" and 'the shift whereby Kepler superseded Ptolemy, or Einstein Newton', there is an important difference in revisability between, on the one hand, theoretical statements of natural science, such as those Kepler, Ptolemy, *etc.*, were concerned with, and, on the other hand, logically or analytically true statements, or statements provable solely on the grounds of meaning. Recall what the revisability claim is: it is that even when a statement such as "α or not α" is once proved solely on the grounds of its meaning and independently of non-semantic matters of fact, still this does not make it 'immune to revision'; its evaluation or reëvaluation is still subject to theory change elsewhere in natural science. But, as I have already shown, theory change in natural science has no effect on the semantic identity, or meaning, of statements and terms. Natural science seeks to determine the *real* essence of certain aspects of the environment; in other words, the *nomological* identity of the *extensions* or *truth-conditions* of symbols, not the symbols' *logical* or *semantic* identity. It follows that once a statements is proved solely from its meaning and independently of non-semantic matters of fact, vicissitudes of theory change in natural science, and reëvaluations of other theoretical statements in natural science, cannot change the statement's truth-value. One could 'revise' the statement only at the cost of a semantic revision, or change of meaning; but then one would merely exchange the statement for a different one of identical syntactic form.

There is much analytic reasoning commonly done in formal logic. In some branches of logic, such as with truth-trees in quantification theory, statements are proved solely from the meanings of their constituent terms. Yet the holists do not take this as showing that genuinely analytic statements exist. They have a good reason for this: formal logic (as it is known today) relies more or less explicitly but nonetheless entirely on the extensional and truth-conditional theory of meaning, which is *certainly* false. Thus formal logic appears to be a contrivance people make to suit their ends; to put it in Quine's words, 'a man-made fabric', subject to convention; and, as such, it gives us no evidence that there are genuinely analytic statements — statements provable solely on the grounds of their own meaning and independently of other statements. In this respect, I fully agree with the holists: the formal logic we have got from Frege and Russell, and which in various guises rests on the extensional-*cum*-truth-conditional theory of meaning, is no evidence for analyticity. But in the face of this, one can either go with the holists, saying that all meaning is but a behavioural convention; or one can reject the Fregean-Russellian logic, with its subsequent developments, as deeply misfounded, and seek the correct foundations elsewhere. From the classical point of view, there is no doubt that the foundations of logic and analytic reasoning in general must be sought in the constitution and functioning of the mind. There is a misconception among many Analytic Philosophers — Quine, most notably — that we already know what the classical point of view is, and what is wrong with it: we are now progressing to higher and better things. But philosophy, need I say, is not like that; many insights which were once well understood become lost and forgotten, and when rediscovered appear new and unusual, having to fight their way to the light of day as though they were never there.

5.3 The Route from Behaviourism to Holism

The route from verificationism to holism showed that, for Quine, statements taken one-by-one cannot be bearers of meaning, and that only symbols in larger linguistic contexts can be said to be meaningful; none of this told us, however, what according to Quine meaning consists in. This issue is the subject of the present section. We shall begin with some expository quotations:

> Meanings are, first and foremost, meanings of language. Language is a social art which we all acquire on the evidence solely of other people's overt behavior under publicly recognizable circumstances. Meanings, therefore, those very models of mental entities, end up as grist for the behaviourist's mill ... (Quine 1969b: 26–27)

The sentence "meanings ... end up as grist for the behaviorist's mill" seems to suggest that Quine proposes semantic *behaviourism*. This is misleading, so long as one regards behaviourism as a *reductive* doctrine requiring that the meaning of an expression be construed in terms of a certain *type* of use of the expression *vis-à-vis* a certain *type* of (factual and counter-factual) stimulatory experiences or conditions occasioning the experiences; that is, so long as one regards behaviourism as requiring that both the correct use of an expression and the stimulatory conditions occasioning the correct use must be bound by some commonality, or property with respect to which all instances of the correct use are identical. We shall see that Quine's account of meaning is not reductive in this sense; it is based on resemblance rather than identity of use and stimulations. However, it is a *behaviour-based* account; which is to say, it holds that the meaning of an expression (word, statements, theory, *etc.*) consists in its overt use and application *vis-à-vis* publicly observable, external stimulations.

One might wonder whether the language-learning argument just quoted — that semantics *must* be behaviour-based since 'language is a social art which we all acquire on the evidence solely of other people's overt behavior' — is a preamble to something more substantial; for it is so thin. But no, the argument from language acquisition is what we may properly regard as Quine's *characteristic* argument, which is rephrased on various occasions as the sole basis for the claim that semantics must be behaviour-based:

> Language is a social art, acquired on evidence of social usage... The learning not just of language but of anything ... is of course external in one respect: it is a conditioning of responses to external stimulation... But the learning of language, in particular, is an external affair also in a further respect, because of the social character of language. It is not just that we learn language by a conditioning of overt responses to external stimulation. The more special point is that verbal behavior is determined by what people can observe of one another's responses to what people can observe of one another's external stimulations. In learning language, all of us, from babyhood up, are amateur students of behavior, and, simultaneously, subjects of amateur studies of behavior... [E]ven those who have not embraced behaviorism as a philosophy are obliged to adhere to behavioristic method within certain scientific pursuits; and language theory is such a pursuit. A scientist of language is ... a behaviorist ex officio. (Quine 1970: 3–4)

Again, some thirty years later:

> In psychology one may or may not be a behaviorist, but in linguistics one has no choice. Each of us learns his language by observing other people's verbal behavior and having his own faltering verbal behavior observed and reinforced or corrected by others. We depend strictly on overt behavior in observable situations. As long as our command of our language fits all external checkpoints, where our utterance or our reaction to someone's utterance can be appraised in the light of some shared situation, so long all is well. Our mental life between checkpoints is indifferent to our rating as a master of the language... There is nothing

> in linguistic meaning, then, beyond what is to be gleaned from overt behavior in observable circumstances. (1987: 5)

Having thus 'embraced behaviourism as a philosophy', Quine proceeds to argue that *reductive* behaviourism — the doctrine that the meaning of an expression is *uniquely* determined by the *type* of factual and counter-factual use of the expression *vis-à-vis* a certain *type* of factual and counter-factual stimulatory experiences — cannot be correct; and hence that, in general, *no* reductive theory of meaning, no theory aiming to identify semantic *properties* of expressions, can be correct. Specifically, Quine's claim against reductive behaviourism is that:

> ... two men could be just alike in all their dispositions to verbal behavior under all possible sensory stimulations, and yet the meanings or ideas expressed in their identically triggered and identically sounded utterances could diverge radically, for the two men, in a wide range of cases. (1960: 26)

Thus formulated, however, the claim appears to contravene his argument for behaviour-based semantics; for, according to that argument, the meaning of an utterance consists in the overt use of the utterance *vis-à-vis* certain triggering experiences. If so, how could the 'identically triggered' and 'identically sounded' utterances he imagines diverge radically in their meaning?

To show that this is possible, Quine deploys an additional premiss in his case against reductive behaviourism: namely, the premiss that "...meaning, supposedly, is what a sentence shares with its translation" (1960: 32). This premiss, in contrast to the premiss that semantics must be behaviour-based, which is backed up by the argument from language learning, is best viewed as a mere supposition; Quine simply assumes it as self-evident. Given the additional premiss, Quine's claim against reductive behaviourism is then reformulated thus:

> The thesis is ... this: manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another. In countless places they will diverge in giving, as their respective translations of a sentence of the one language, sentences of the other language which stand to each other in no plausible sort of equivalence however loose. (1960: 27)

The claim is that the sameness of dispositions to verbal behaviour (the sameness of utterances that *would* be made in response to the same stimulations) is compatible with a multiplicity of translation schemes; and hence, since 'meaning is what a sentence shares with its translation', that semantic behaviourism — despite being the only candidate for a reductive semantical theory passing Quine's argument from language acquisition —

is false; therefore behavioural holism is the only alternative. His argument for this claim begins thus.

Consider the following *gedankenexperiment*. A linguist is engaged in translating, without any help from interpreters, the language of an alien culture. Using Quine's own terminology, the linguist is engaged in "*radical* translation, i.e., translation of the language of a hitherto untouched people" (1960: 28). Suppose the linguist correctly identifies the native signals or expressions for assent and dissent, being therefore able to collect inductive evidence for the correct translation of other simple expressions of the language. For example, assume the people often say "Gavagai" in the presence of rabbits; the linguist surmises that "Gavagai" means the same as (is synonymous with) "Rabbit", and proceeds to gather evidence in support of this hypothesis by yes/no queries under various stimulatory circumstances. Finally, suppose the linguist works with the following semantical theory. A native expression is *synonymous* with an English expression just in case the two expressions have the same *stimulus meaning*; that is, just in case a native speaker *would* assent to (/dissent from) a yes/no query deploying the native expression under *those and only those* stimulatory conditions that *would* prompt an English speaker to assent to (/dissent from) a yes/no query deploying the English expression.

There are two salient features one should note concerning this scenario. Firstly, the stimulus-meaning theory conforms to the requirements of Quine's characteristic argument from language learning; more precisely, the theory is a deliberately simplified version of *reductive behaviourism*. Secondly, the scenario of radical translation is one wherein the consequences of reductive behaviourism *and* the premiss that 'meaning is what a sentence shares with its translation' can be readily investigated. Quine uses the stimulus-meaning theory — which, so he takes it, would have to be adopted by a semanticist engaged in radical translation — not as a full-blown semantical theory, but to demonstrate that, even under simple conditions, the sameness of dispositions to verbal behaviour under-determines the sameness of meaning: in Quine's words, that 'two men could be just alike in all their dispositions to verbal behavior under all possible sensory stimulations, and yet the meanings expressed in their identically triggered and identically sounded utterances could diverge radically, for the two men, in a wide range of cases'.

Quine puts forth, in order to demonstrate his thesis, two sorts of argument based on the scenario of radical translation. Firstly:

5.3.1 The argument from collateral information.

Take the native expression "Gavagai", and suppose it does refer to rabbits, so that "Gavagai" and "Rabbit" are, as a matter of fact, co-extensional. However, assuming the stimulus-meaning theory (which is a version of behaviourism, which is the only *reductive* account allowed by the language-learning argument), the co-extensionality of "Gavagai" and "Rabbit" will not be sufficient for their stimulus-synonymy; *and*, more importantly, the stimulus-meaning of "Gavagai" for a native speaker will never be identical to the stimulus-meaning of "Rabbit" for an English speaker. The reason is that the stimulus-meaning of an expression depends on what Quine calls "prior collateral information"; that is, information subsidiary to the usual prompting stimulus associated with the expression. For example, the stimulus-meaning of "Gavagai" for a native speaker will depend on the speaker's background knowledge of the local habitat of rabbits, enabling the speaker to assent to (/dissent from) such yes/no queries as "Gavagai?" on the evidence of "nothing better than an ill-glimpsed movement in the grass..." (1960: 37). Such collateral information will not be included in the stimulus-meaning of "Rabbit" for an English speaker, or the linguist. Hence although "Gavagai" and "Rabbit" have identical extensions, their stimulus-meanings will never be identical, and the one should not be translated as the other, according to semantic behaviourism.

This argument does not directly address Quine's claim that 'two men could be just alike in all their dispositions to verbal behavior under all possible sensory stimulations, and yet the meanings expressed in their identically triggered and identically sounded utterances could diverge radically, for the two men, in a wide range of cases'. Rather, the argument shows that even when the stimulatory experiences are the same or as alike as one wishes for the native speaker and the linguist, still the stimulus-meanings of their (co-extensive) expressions differ because of the variance in their collateral information; and that consequently translation is under-determined by empirical data (assuming semantic behaviourism). Yet 'meaning is what a sentence shares with its translation'; so semantic behaviourism must be false; in general, any *reductive* theory of meaning must be false, according to Quine.

It is remarkable that Quine notices that stimulus-meaning falls short of one's intuitive conception of meaning:

> ... stimulus meaning as defined falls short in various ways of one's intuitive demands on "meaning" as undefined... Yet stimulus meaning ... may be properly looked upon still as the objective reality that the linguist has to probe when he undertakes radical translation... We do best to revise not the notion of stimulus meaning, but only what we represent the linguistic as doing with stimulus meanings. The fact is that he translates not by identity of stimulus meanings, but by significant approximation of stimulus meanings. If he translates 'Gavagai' as 'Rabbit' despite the discrepancies in stimulus meaning imagined above, he does so because the stimulus meanings seem to coincide to an overwhelming degree and the discrepancies, so far as he finds them, seem best explained away or dismissed as effects of unidentified interferences. (1960: 39-40)

'Stimulus meaning as defined' is the *reductionist* meaning assumed by behaviourism; in contrast, '"meaning" as undefined' is 'whatever a sentence shares with its translation'. Since there is a conflict between the two, and since, as he is convinced, neither the premiss that semantics must be behaviour-based nor the premiss that meaning is what is preserved in translation is at fault, Quine decides to abandon semantic reductionism. His attempt to salvage what he can of meaning is to insist that translation proceeds 'not by identity of stimulus meanings, but by significant approximation of stimulus meanings'.

The second argument directly addresses the thesis that sameness of dispositions to verbal behaviour is compatible with a variety of mutually incompatible translation manuals.

5.3.2 The argument from inscrutability of reference.

Consider again the expression "Gavagai", and suppose, not that "Gavagai" and "Rabbit" are co-extensive but that they are *stimulus-synonymous* (notwithstanding the argument from collateral information); that is, that a native speaker would assent to (/dissent from) "Gavagai?" under those and only those stimulatory conditions that would prompt an English speaker to assent to (/dissent from) "Rabbit?". Does the stimulus synonymy of "Gavagai" and "Rabbit" *guarantee the co-extensiveness* of the general terms "gavagai" and "rabbit"? Quine argues that it does not, and hence that the sameness of 'stimulus meaning as defined' under-determines the sameness of 'meaning as undefined'; in other words, that the *reductionist* meaning assumed by semantic behaviourism under-determines translation:

> ... consider 'gavagai'. Who knows but what the objects to which this term applies are not rabbits after all, but mere stages, or brief temporal segments, of rabbits? In either event the stimulus situations that prompt assent to 'Gavagai' would be the same as for 'Rabbit'. Or perhaps the objects to which 'gavagai' applies are all and sundry undetached parts of rabbits; again the stimulus meaning would register no difference. When from the sameness of stimulus meanings of 'Gavagai' and 'Rabbit' the linguist leaps to the conclusion that a gavagai is a whole enduring rabbit, he is just taking for granted that the native is enough like us to have a brief general term for rabbits and no brief general term for rabbit stages or parts... We could equate a native expression with any of the disparate English terms 'rabbit', 'rabbit stage', 'undetached rabbit part', etc., and still, by compensatorily juggling the translation of numerical identity and associated particles, preserve conformity to stimulus meanings... (1960: 51-54)

Thus the sameness of dispositions to verbal behaviour is compatible with a variety of mutually incompatible translation manuals, and hence — given that 'meaning is what a sentence shares with its translation' — semantic behaviourism must be false. Yet, Quine insists, semantics cannot but rest on linguistic behaviour, and it cannot but uphold the principle that meanings are whatever is preserved in translation. He resolves the dilemma, as we know, by stipulating that translation proceeds not by identity of behavioural meanings but by approximation.

What is, then, Quine's semantic holism? It is a *behaviour-based* account of meaning, but it is not to be confused with reductive behaviourism. According to behavioural holism, but contrary to reductive behaviourism, linguistic expressions have no *determinate* semantic identity; meanings are not scientifically individuable properties of expressions. Correspondingly, meanings *under-determine* their extensions; they are not the Conservative extension-determiners and truth-condition-determiners of expressions.

Further, meanings are what an expression shares with its *translation*, and translation cannot proceed severally, term-by-term or sentence-by-sentence. This is because it must take into account, or sometimes disregard, *collateral information*; and it must take into account or sometimes disregard that meanings *under-determine reference*; and that there could be other equally good yet incompatible translation schemes. Thus whatever fuzzy 'semantic identity' an expression might be said to have, that identity is bound to depend on the indeterminate 'semantic identity' of its larger linguistic and behavioural context: words, sentences, theories, *etc.*, are meaningful not one-by-one but only insofar as they have a role within larger contexts; and the primary bearers of meaning are the contexts themselves, not individual expressions.

Lastly, since translation is so under-determined on all sides, and since there is nothing anyone could do about it, Quine pleads for 'semantic tolerance', or *charity of interpretation* (see (1960: 59, 69)). (The term "charity of interpretation" has been much appropriated by Putnam, *et al.* Some readers may find it interesting, though, that the term is due not to Quine, let alone Putnam, but to Locke, in an altogether different semantics; see (1975: III, IX, 22).)

5.3.3 Four refutations of behavioural holism.

Quine's argument against reductive behaviourism and for behavioural holism rests on the premiss that semantics must be behaviour-based, backed by the language-learning argument, and on the premiss that 'meaning is what a sentence shares with its translation', assumed without any argument; hence Quine proceeds, setting up the *gedankenexperiment* of radical translation, by the arguments from collateral information and inscrutability of reference, to the conclusion that behaviour-based translation is under-determined by its evidence, and to semantic holism. We can therefore distinguish four targets of counter-argument: *(i)* the premiss that semantics must be behaviour-based; *(ii)* the premiss that 'meaning is what a sentence shares with its translation'; *(iii)* the argument from collateral information; *(iv)* the argument from inscrutability of reference. If either either of the premisses fails, or both of the inferences fail, Quine's route from behaviourism to holism will be closed; and since the routes from verificationism to holism, and from the various minor alternatives of Plato's beard to holism, have already been closed, his overall argument for semantic holism will be refuted.

5.3.3.1 The language-learning argument.

The gist of the argument is that semantics must be behaviour-based, in that the meaning of an expression can consist only in the expression's overt use and application with respect to factual and counter-factual publicly observable stimulatory conditions, since one's knowledge of the meaning is acquired solely on the evidence of the expression's use and application with respect to such inter-subjectively available stimulations.

This is a surprisingly bad argument. Even if one's knowledge of the meaning of an expression were acquired *solely* on the evidence of the expression's use and application with respect to certain stimulatory circumstances (which is unlikely, for the mind probably makes use of its already available resources to learn the meaning of a new expression), still it would not follow that the meaning of the expression must consist in the *evidence* for the knowledge acquisition. The way one acquires knowledge of the semantic identity of a symbol is largely irrelevant to the symbol's semantic identity. In general, the way one acquires knowledge of the identity of whatever entity in whatever respect is irrelevant to the identity of the entity in that respect. For example, it may be that one learns that something is gold by observing that it is soluble in *aqua regia*; still, being gold is not the same property as being soluble in *aqua regia*. In the linguistic case, it may be that one learns the meaning of an expression by observing under what stimulatory circumstances the expression is used and applied; but it does not follow that the expression's meaning consists in its use *vis-à-vis* such circumstances.

Perhaps, yielding to Quine's plea for charity in interpretation, the implicit reasoning behind his argument is this: one often gets to know the meaning of a word in a learning situation by ostension; and if what one gets to know (that is, the meaning) is not *what* the object or set of objects or kind ostended *is* — which it is not, since the extensional theory of meaning is certainly false — it must be that what one gets to know is nothing but *how* to use and apply the word in actual and hypothetical situations.

This is quite a progress on the extensional theory; for the extensional theory is false simply on the grounds of what it says (it is necessarily false), whereas the behavioural account seems to make a genuine claim about what meaning consists in. However, even so charitably interpreted the argument does not go through. One difficulty with the behavioural account is that verbal use vastly outruns the stimulatory conditions of any learning situation. The use of a word, as Descartes was one of the first to note (in Part V of the *Discourse on Method*), is not just a matter of responding to stimulatory conditions similar to those in which the word was learned; rather, words are general instruments which can and are used, once acquired, wholly independently of any specific stimulatory circumstances.

But by far the greatest worry is that, with his argument for behaviour-based meaning, Quine screens off the human mind from what we are and how we mean. Yet the mind gives us an excellent solution to the Cartesian universality and *stimulus-independence of verbal use*; it allows us, in learning the meaning of a new word, to learn not merely *how* to use the word in response to such and such stimulatory conditions, but rather *that* the word represents a certain mind-dependent, nominal property; and this in turn allows us to use the word, once learned, regardless of the original learning situation. The language-learning argument, which Quine characteristically relies on to establish his premiss that semantics must be behaviour-based, certainly does not establish the premiss; and, without it, Quine's conclusion of behavioural holism is much like the emperor's new clothes: not to see through it we would have to be, as Quine takes us, unconscious.

5.3.3.2 The premiss that meaning is what a sentence shares with its translation.

This premiss effectively stands the business of translation on its head: for it rules, by decree, that meaning is what a sentence shares — or what is preserved — in translation, rather than that the goal of translation is to preserve as much of meaning as possible. The latter, not the former, construal of translation is no doubt correct, as every bilingual and every translator can attest. Whether one is a behaviourist or mentalist, it will turn out that each language and linguistic community generates its own meanings. In the case of mentalism, there is a class of semantically simple mental symbols, and these are universal for every speaker who acquires them; but complex symbols will be products of the speakers themselves, and as such will differ from community to community, and to some extent from person to person and time to time in each community. In general, meanings, whether of an individual speaker or, by a sort of social regimentation, of a linguistic community, are made within the socio-linguistic environment of the community; *not across such communities*. So when it comes to translating between languages, it will not be practically possible to find a one-to-one mapping or correspondence of meanings, and any rendition of complex expressions of the one language in the other will have to be a matter of pragmatic approximation. But then, meanings cannot be preserved in translation, and cannot be identified with 'what a sentence shares with its translation'.

The problem of translation does not properly belong to semantics, even if semantics is construed as behavioural rather than mentalistic; it is a problem of rendering the meanings of one language intelligible in another, and as such it belongs to the pragmatics of language. It is therefore a mistake to constrain the semantic problem of telling the nature of meaning by the pragmatic problem of translation.

5.3.3.3 The argument from collateral information.

The gist of the argument is that the behavioural meanings of a native and a linguist compiling a translation-manual can never be the same, and the linguist can never quite understand the native's meaning, because verbal dispositions cannot be quite identical for the native and the linguist; for instance, the native's use of "Gavagai" will always differ from the linguist's use of "Rabbit", *etc.*; in turn, this is because verbal dispositions depend on such prior collateral information as the native's knowledge of the local habitat of rabbits, and so forth.

Notice that Quine helps himself, as it suits him, to such meaning-connoting terms as "information" and "knowledge". But we may suppose that information and knowledge are to be likewise explained by means of verbal dispositions. If so, then Quine's argument from collateral information amounts to saying no more than that the native's dispositions to use, say, "Gavagai" are never the same as the linguist's dispositions to use any word of English, "Rabbit" among other; so the linguist's decision to translate "Gavagai" as "Rabbit" must be a matter of balance of the overall translation scheme; and hence behavioural holism as a theory of meaning clearly follows, *given Quine's two premisses*.

There is no question that, given the premiss that semantics must be behaviour-based and the premiss that meaning is what is preserved in translation, behavioural holism follows; I will grant this without ado. But neither of the premisses hold, which undermines the conclusion more than enough to be rejected. Surprisingly, although so much attention has been paid to the argument from collateral information (and even more to the argument from inscrutability of reference), the premisses have usually passed without so much as being mentioned.

Still, it will be instructive to look at what happens to the conclusion of semantic holism if we remove one or the other, or both of the premisses. Keeping the premiss that meaning is what a sentence shares with its translation seems to be sufficient for holism of some sort, whether behaviour-based or otherwise. For if one identifies the meaning of a symbol not *within* the socio-linguistic environment to which the symbol belongs, but *across* such environments, making its identity dependent on the aims of translators who try to match disparate languages, then the meaning will be fixed only by an overall pragmatic balance of a translation scheme. But this is to smother semantics at its inception.

It is obvious that keeping the premiss that semantics must be behaviour-based, whilst rejecting the translation premiss, defeats holism; for *reductive* behaviourism remains an option, although not a very live one. Perhaps the holist would object that suspending the translation premiss will not save behaviour-based meaning from turning holistic, since meaning has to be *inter-subjective* and communal, so that communication even within the same language requires 'translation' of a sort. But no; the reductionist, behaviourist or mentalist, can and should insist that meaning is primarily a *personal* matter. The *behaviourist* might say that it is fundamentally an *individual* speaker who means so and so by using or tokening a word, and that the semantic identity of the word is fixed by the speaker's *personal* dispositions to use it; communication depends on resemblance of individual meanings, but it does not thereby destroy the reductive identity of meanings. Analogously, the *mentalist* can and should say that if, *e.g.*, Jane uses the word "gold" to stand for or express the idea **yellow precious metal**, then what she means or denotes by "gold" is the nominal essence [yellow precious metal]; and if John uses "gold" to express **yellow malleable metal**, then what he means or denotes is [yellow malleable metal]; but they can still communicate since, as it might be, **yellow metal** is enough of a commonality for their purposes. Suppose the holist still objects that regarding meaning as a personal matter will not save it from holism, since reductionist meaning requires a uniform identity *over time*, which, as a matter of fact, meaning does not have; even conversing with oneself or recalling over time is a 'translation' of a sort. But again, the reductionist, behaviourist or mentalist, will insist that although there is a finite number of semantic simples with a uniform identity over time (and these the organism can neither make nor change), *complex meanings can be variable over time*; the behaviourist would say that the semantic identity of a word, as used on a certain occasion by a certain speaker, is fixed by the speaker's dispositions to use the word *on that occasion*; the mentalist would say that the word's semantic identity is fixed by the semantic identity of the idea, *qua token* of a mental symbol, which the speaker uses the word to express *on that occasion*, and the semantic identity of the idea consists in its denoting a certain nominal essence. In short, once the premiss that meaning is what is preserved in translation is rejected, in any of its forms, holism will not follow, behavioural or whatever, even granting to Quine all the prior collateral information he ever wished.

Finally, suppose we reject both the premiss that semantics is behaviour-based and the premiss that 'meaning is what a sentence shares with its translation'. What is left of holism? Quine tried to get to holism not only *via* these two premisses, but also *via* verificationism; but we have already seen that that way is closed; and not only *via* verificationism, but also *via* the minor alternatives of Plato's beard; but we have seen that that way too is closed.

5.3.3.4 The argument from inscrutability of reference.

Quine argues that even assuming "Gavagai", as used by a native speaker, is *stimulus-synonymous* to "Rabbit" as used by the linguist — *i.e.*, assuming "Gavagai' and "Rabbit" are semantically identical by the criteria of reductive behaviourism — still, this does not *uniquely* fix the linguist's problem whether to translate "Gavagai" as "Rabbit"; the reason is that behaviour-based meaning under-determines reference, or that reference is inscrutable by behavioural criteria: thus, for example, "Gavagai" could refer to whole rabbits, or temporal rabbit stages, or undetached rabbit parts, with no difference registered by the behavioural criteria.

Allowing ourselves to speak of properties (contrary to Quine's nominalism), we might express the argument thus: the properties of being a rabbit, being a temporal rabbit stage, or being an undetached rabbit part, are always instantiated simultaneously; so the native's dispositions to apply "Gavagai" to rabbits, to temporal rabbit stages, to undetached rabbit parts, will be identical; the linguist might be disposed to apply "Rabbit" only to whole enduring rabbits, but whether "Gavagai" should be translated as "Rabbit" remains an open question; it might be equally well translated as "temporal rabbit stage" or "undetached rabbit part". In other words, *translation is indeterminate* since *reference is under-determined* by verbal dispositions; so reductive behaviourism — and reductionism in general — must be false.

There are two points I wish to make. Firstly, the gist of Quine's case is that meaning under-determines reference, or — to put it in my nomenclature — that meaning is not an extension-determiner. I agree; in fact, one of my polemical theses is that the Conservative principle of semantic individuation, that meaning fixes extension or truth-condition, must be rejected. But Quine, along with most or all of Analytic Philosophy, assumes that semantic *reductionism,* whether behaviourism or mentalism, requires that principle; that reductionism stands and falls with the rule of extension-meaning supervenience. That this is not so insofar as mentalism is concerned I have already made clear; the key is that the meaning of a symbol token (mental or public) consists in its *denoting* a unique *nominal* property, and as such the meaning does not consist in or require *referring* to a unique extension.

Secondly, Quine argues that behavioural criteria — as used by the linguist in the scenario of radical translation — are not enough to tell whether "Gavagai" (or, for that matter, "Rabbit") refers to whole enduring rabbits, temporal rabbit stages, or undetached rabbit parts. The issue is not merely whether behaviour-based meaning is an extension-determiner; we may take it as read that it is not. The issue is rather whether sole behavioural criteria are sufficient to distinguish between reference to such grossly different particulars as a whole enduring rabbit, a temporal rabbit stage, or an undetached rabbit part. Quine argues that they are not, and again I fully agree. But this does not show that we have to accept behavioural holism; it shows only that behaviourism, reductive or holistic, is false. Mentalism has no difficulty with distinguishing between *reference* to whole enduring rabbits, temporal rabbit stages, or undetached rabbit parts. According to mentalism, the word "rabbit" is used, by a speaker on a certain occasion, to stand for an idea in the speaker's mind on that occasion, say, **a burrowing herbivorous rodent**. This is not sufficient to determine the extension {rabbit}, or the natural kind <rabbit>; but it is sufficient to determine the nominal property [a burrowing herbivorous rodent], and that is what the speaker *means* (has in mind), or *denotes*, when tokening "rabbit" on that occasion. Suppose another speaker uses "rabbit" to stand for a different idea, say, **an undetached tail of a burrowing herbivorous rodent**. If the speakers' ideas differed only in some minor detail — as in practice most *complex* ideas will differ in detail from person to person and, for each person, from time to time — the speakers might have a difficulty sorting out just where their ideas differ; but then, minor differences would not impair communication, so they would be justified in disregarding them. But in the case of a major difference, the size and significance of **an undetached tail**, they should be able to sort out their disagreement. Clearly, they could not sort it out solely on the evidence of their overt use of the word "rabbit" with respect to particular rabbits; for rabbits partake of the nominal property [a burrowing herbivorous rodent] just in case they partake of [an undetached tail of a burrowing herbivorous rodent]; thus the speakers' dispositions to use "rabbit" would be much the same. But, not being behaviourists, they might contrive to tell each other what constituent ideas the *mental descriptions* they use the word "rabbit" to stand for comprise. It would be practically impossible for them to spell out their mental descriptions in terms of *semantically simple* ideas; for one thing, any complex idea of rabbits or undetached rabbit tails would be very complex indeed; for another, their public language probably would not have words for each type of simple idea. But in terms of *categories* of simple ideas, and in terms of constituent complex ideas for which they already have settled words, they should be able to spell out their ideas with a bit of introspective labour. When they have succeeded, they might perhaps decide to agree on a *common acceptation* of the word "rabbit", adjusting their ideas accordingly; and although their common acceptation would not be sufficient to determine the extension {rabbit}, or the natural kind <rabbit>, it would be enough to distinguish, when they *refer* to particular objects in their environment, between whole enduring rabbits, temporal rabbit stages, undetached rabbit parts, and so forth. That, as a matter of fact, people in each linguistic community do succeed in regimenting their use of words so as to observe such gross distinctions in what they refer to, is a powerful evidence against either reductive or non-reductive behaviourism, and in favour of semantic mentalism. For people could not regiment the use of their words to the extent they in fact do, unless they had minds comprising a representational code which they are able to express by means of their public languages; mind-less, behaviour-based meaning, as Quine himself convincingly shows, could not do it.

5.4 Remarks on Middlebrow Pragmatism

It might seem, given our survey of the holistic-*cum*-behavioural grounds, that no one would want to build the family house on such sifting sand; yet that is where most people build today. In this final section, I will show briefly how Quine's holism has been adopted and adapted by one prominent exponent of the Middlebrow strand in Analytic Philosophy, Hilary Putnam. We shall see that Putnam entirely accepts Quine's semantic holism and his arguments for it; in addition, he brings forth certain implicit aspects of Quine's position: the *coherence theory of truth* and the *satisfaction theory of reference* which comport with holism. These theories, together with the Conservative doctrines of the division of semantic labour and the contribution of environment, comprise the core of what Putnam calls "internal" or "pragmatic realism".

5.4.1 Meaning holism.

Putnam (1988) accepts without reservation Quine's (1951) argument from confirmation holism to semantic holism:

> If the sentences of which a theory consists had their own independent experiential meanings, or made so many separately testable claims as to what experience will be like, then one could test a scientific theory by testing sentence 1 and testing sentence 2 and testing sentence 3 and so on. But, in fact, the individual postulates of a theory generally have no (or very few) experiential consequences when we consider them in isolation from the other statements of the theory... As Quine puts it, sentences meet the test of experience "as a corporate body," and not one by one. (Hence the term "holism.") (1988: 8-9)

Hence Putnam concludes that "[i]f language describes experience, it does so as a network, not sentence by sentence" (*ibid.*: 9), and that meaning holism must be true.

Similarly, Putnam accepts Quine's premisses that semantics must be behaviour-based, and that 'meaning is what a sentence shares with its translation'. For example, "... meaning, we should recall, if it is anything, is what we try to preserve in translation" (*ibid.*: 29); and, to prevent any misunderstanding about it:

> ... why should we have such a notion as meaning at all? But this is not really such a puzzle: the best way to get along with people who speak a different language — or, on occasion, even to get along with people who speak one's "own" language in a different way — is to find an "equivalence" between the languages such that one can expect that — after due allowance for differences in beliefs and desires — uttering an utterance in the other language in a given context normally evokes responses similar to the responses one would expect if one had been in

> one's own speech community and had uttered the "equivalent" utterance in one's own language. (1988: 25–26; *op. cit.*, Section 2.2)

In other words, the meaningfulness of an utterance consists in its overt dispositional use under publicly observable circumstances; and the semantic equivalence between utterances of different languages (or utterances of the same language on different occasions) consists in the *similarity* of their overt use in their respective environments or circumstances. Likewise, Putnam endorses Quine's arguments from the indeterminacy of translation (1978a: 44–45), and his dissociation of the theories of meaning and reference (1978b: 97–100; 1988: 38). Finally, like Quine, Putnam emphasises the role of charity in interpretation in telling the identity of a meaning (1988: 13–15). In a sum, Putnam's account of meaning closely follows Quine's holism; the same arguments I have brought against Quine in this and the last chapter apply equally here.

5.4.2 The coherence theory of truth.

The meaning of a sentence, according to semantic holism, is dependent on the entire conceptual scheme the sentence belongs to; Putnam (1978c; 1981; 1988) claims that the same kind of *conceptual relativity* affects also the truth of the sentence. Truth is indeed *objective*, Putnam says, in that "it is a property of truth that whether a sentence is true is logically independent of whether a majority of the members of the culture *believe* it to be true" (1988: 109); however, this is not to say that truth is "independent of what human beings know or could find out..." (*ibid.*); truth is not only objective but also relative to one's conceptual scheme. What does the conceptual-scheme relativity of truth consist in? Consider the problem of distinguishing *fact* from *convention* in determining the truth-value of a sentence. Putnam takes it that Quine (1951) has shown conclusively that "the very distinction between 'fact' and 'convention' ... collapses when construed as a sharp dichotomy" (1988: 112), and that consequently the determination of the truth-value of a sentence is, at least in part, a matter of convention. It follows, granted there is a "diffuse background of empirical facts" (*ibid.*: 113), that the truth-value of a sentence cannot be determined unless within a conceptual scheme which the sentence belongs to; in other words, truth cannot be judged absolutely, independently of conceptual schemes: it can be judged only within, or relative to, a conceptual scheme.

What this argument shows, Putnam says, is not that each conceptual scheme divides *the* world, *the* reality, in its own way, applying its own relativistic conception of truth with respect to what there is; for that would be to assume that there is a unique way the world is, and that some conceptual schemes may be more true of the world than other. The argument shows, according to Putnam, that though "[w]e can and should insist that some facts are there to be discovered and not legislated by us ... this is something to be said when one has adopted a way of speaking, a language, a "conceptual scheme." To talk of "facts" without specifying the language to be used is to talk of nothing" (1988: 114).

Given the conceptual-scheme relativity of truth, Putnam claims that truth consists in "...some sort of (idealized) rational acceptability — some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system* — and not correspondence with mind-independent or discourse-independent 'states of affairs'" (1981: 49–50); hence the *coherence* theory of truth, or rational acceptability. (Putnam prefers to avoid the term "coherence theory of truth", because of its other uses; this, however, need not concern us here.) Putnam (1988) further elaborates on the account thus:

> ... *a statement is true of a situation just in case it would be correct to use the words of which the statement consists in that way in describing the situation.* Provided the concepts in question are not themselves ones which we ought to reject for one reason or another, we can explain what "correct to use the words of which the statement consists in that way" means by saying that it means nothing more nor less than that a sufficiently well placed speaker who used the words in that way would be fully warranted in *counting* the statement as true of that situation. (1988: 115)

Putnam adds that he is not being *cute* in offering this 'explanation'. The account is intended to be *non-reductive*: truth is explained in terms of 'correct use', and 'correct use' in terms of 'warranted use'. It comports well with semantic holism; in effect, the account *is* semantic holism, with "use" replaced by "correct use" or "warranted use".

Everything that counts against semantic holism therefore counts against the coherence theory of truth; and that is quite a lot. Here I will pick on a fault which is not related directly to semantic holism. To begin with, I concede that Putnam's premiss that truth has to be judged relative to, or from within, one's conceptual scheme, so that it is conceptual-scheme relative, is correct. But I do not think that the conceptual-scheme relativity of truth is anything like a deep fact about truth. In my view, truth-values are epistemic values of *tokens* of *sentences*; that is, they are values that sentence-tokens bear because they are epistemically evaluable, at least in principle, in certain ways. It follows that truth or falsehood cannot but depend on what sentences are available in the linguistic or, in general, representational system one uses. So it makes hardly any sense to speak of truth or falsehood independently of any conceptual or symbolic scheme. Similarly, Putnam is quite right in saying that what facts we recognise, what facts there are for us, cannot but depend on what conceptual scheme we use; there is no point in trying to grasp what facts there exist 'absolutely', or independently of any symbolic system whatever.

However, Putnam takes it that the thesis of the conceptual-scheme relativity of truth has the following metaphysical import: that it makes no sense to speak of the way the world, the reality, is; that one cannot sensibly assume that there is a unique way the world is, and that some conceptual schemes may represent the world more accurately or fully than others. This would follow only if the alternative picture, that there is a unique way the world is, were incompatible with the conceptual-scheme relativity of truth. But we can clearly acknowledge that truth is conceptual-scheme relative, *and* hold that there is unique way the world is, that one's conceptual scheme succeeds in representing some aspects of the world whilst failing to represent others, and that some conceptual systems succeed in representing the environment more fully than others. Accordingly, we can acknowledge that what facts we recognise, what facts there are for us, depends on what representational system we use, *and* hold that there is a unique way the world is, *etc*. The thesis of the conceptual-scheme relativity of truth and facts-for-us is correct regardless of one's background ontology, and gives no special support to the coherence theory of truth as opposed to some version of the correspondence theory; all the thesis says is that we can grasp what facts there are, and hence what statements are true, only from within one representational system or another. But then, truth can still be a matter of correspondence to facts, notwithstanding Putnam's bleensome tale of an objective but in itself fact-less world, a world with no properties.

5.4.3 The satisfaction theory of reference.

We have seen that, according to semantic holism, the meaning of an expression depends on the entire conceptual scheme the expression belongs to; and that the same conceptual-scheme relativity pertains to the truth of statements. Putnam (1978c; 1981; 1988) argues that, similarly, the reference of expressions is conceptual-scheme relative:

> In an internalist view also, signs do not intrinsically correspond to objects, independently of how those signs are employed and by whom. But a sign that is actually employed in a particular way by a particular community of users can correspond to particular objects *within the conceptual scheme of those users*. 'Objects' do not exist independently of conceptual schemes. *We* cut up the world into objects when we introduce one or another scheme of description. Since the objects *and* the signs are alike *internal* to the scheme of description, it is possible to say what matches what... Indeed, it is trivial to say what any word refers to *within* the language the word belongs to, by using the word itself. What does 'rabbit' refer to? Why, to rabbits...! What does 'extraterrestrial' refer to? To extraterrestrials (if there are any)... Of course, the externalist agrees that the extension of 'rabbit' is the set of rabbits and the extension of 'extraterrestrial' is the set of extraterrestrials. But he does not regard such statements as telling us what reference *is*. For him finding out what reference *is*, i.e. what the *nature* of the 'correspondence' between words and things is, is a pressing problem... For me there is little to say about what reference is within a conceptual system other than these tautologies.

> The idea that causal connection is necessary is refuted by the fact that 'extraterrestrial' certainly refers to extraterrestrials whether we have ever causally interacted with any extraterrestrials or not! (Putnam 1981: 52)

So reference, like meaning and truth, 'does not transcend use' (*cf.* Putnam (1988: 115–16)); there is 'little to say' about reference save that "rabbit" refers to rabbits, *etc*. Notice, though, that Putnam's Middlebrow realism includes not only the *internalist*, non-causal, use-based, satisfaction account of reference here outlined, but also, independently, the doctrines of the division of linguistic labour and the contribution of environment; yet these are *externalist*, causal accounts of reference (see (1988: 109; 1975: 245–46)). Remarkably, the discord within Putnam's position has subsequently given rise to, on the one hand, what is currently known as "externalism" in the theory of linguistic and mental content, and on the other hand, the kind of internalism concerning meaning, truth, and reference as described in this section.

Putnam's argument for the satisfaction theory of reference is analogous to that for the coherence theory of truth; *viz.*, that since reference is relative to the conceptual scheme one uses to represent the environment, and since what objects one is able to recognise and refer to also depends on one's conceptual scheme, it follows that reference, like meaning and truth, 'does not transcend use', and that it is trivial: a word refers to what it says it refers to, and there is nothing more to it. The chief reason why the satisfaction theory of reference for public expressions fails is that it relies on and complements behavioural holism in semantics, against which there is much to say. A more specific reason is that the thesis of the conceptual-scheme relativity of reference is true quite independently of the satisfaction theory of reference, and does not require the internalist ontological picture. To say that reference is conceptual-scheme relative is simply to say that reference is a property of the symbols of one or another system of representation (public or mental), and that one cannot refer to objects except by means of one's symbolic system. We may concede to Putnam that 'signs do not intrinsically correspond to objects, independently of how those signs are employed and by whom'; we may agree that 'a sign that is actually employed in a particular way by a particular community of users can correspond to particular objects *within the conceptual scheme of those users*'; but we need not conclude that 'objects do not exist independently of conceptual schemes', and that 'since the objects *and* the signs are alike *internal* to the scheme of description, it is possible — indeed, trivial — to say what matches what'. The reason is that the conceptual-scheme relativity of reference is quite compatible with there being objects (instances of natural kinds, events, *etc.*) in the environment, existing independently of our symbolic systems; and there is no conflict in the view that we succeed in referring to some objects but not in referring to others. To suppose otherwise is to pretend that the way the world is depends on the way we represent it, and even that we can make the world as we will simply by representing it as we will; a kind of sorcery. In the absence of more veridical reasons for accepting the internalist metaphysical picture, the conclusion that the reference of public symbols is merely a matter of trivial satisfaction between them and 'internal objects' does not follow from the true premiss that reference is relative to one's symbolic scheme.

Putnam's trouble is that he so much desires a compromise as to render his position absurd and incoherent. His non-causal, internalist, use-based theory of reference contravenes his causal, externalist account by the division of linguistic labour and by the contribution of environment. Again, in setting up his Twin-Earth scenario to show that the environment does contributes to reference and meaning, Putnam makes it clear that everything on Earth and on Twin Earth is identical; in particular, the use of the word "water" is identical in English and Twin-English. Hence we have it, assuming the satisfaction theory of reference and use-based theory of meaning, that "water" in English and Twin-English are *co-referential* and *synonymous*; assuming the doctrine of the contribution of environment, however, we have it that they are *not co-referential* and *not synonymous*. The Middlebrow view does not pay off in philosophy (though whether it does in making a career of philosophy is another matter).

5.4.4 Holism's wake.

Holism is a form of life with a past. Its origins, as we have seen, are the origins of Analytic Philosophy, when Frege argued that terms are meaningful only in the context of sentences, and Russell argued that it is only the sentential contexts themselves, never terms, which are bearers of meaning. We have been trailing the evolution of holism from Frege and Russell, *via* the semantic verificationism of Carnap, *via* Quine's arguments against Plato's beard, *via* his arguments against verificationism and for holism, against reductive behaviourism and for behavioural holism, till we have reached the Middlebrow shores of contemporary pragmatic realism. It ensues that the abandonment of term-based semantics *en route* to semantic holism has never been well justified. Remarkably, not only none of the arguments against term-based semantics and toward semantic holism succeed, the issue of term-based semantics has hardly been so much as touched upon by the arguments, since the semantical views targeted by them were the various versions of the extensional theory, or misconstrued ideational theory, which are not genuine theories of meaning. Term-based semantics does rely on classical mentalism; but when mentalism is misunderstood in essential aspects, as it was by Frege, Russell, Quine, and much of Analytic Philosophy, it will seem that meaning is not a matter of *having in mind*, but merely a matter of public performance and convention. Once behaviourism is accepted, so that meaning-bearing states come to be regarded as modes of verbal and other behaviour, holism is inevitable; and then, one will tend to think that not only meaning, but also reference, truth, knowledge, one's own identity, and even being as such are no more than a convention. For there will seem to be nothing to us but stimulus-dependence, which cannot afford us an identity in regard of what we mean, let alone who we are; and eventually, there will seem to be nothing to our being — and being as such — but a sort of an unauthorised 'text'. What bearing this neo-sophism has had on Academic and wider culture I leave to the reader's observation. But I may say that what is at stake in the study of semantics is not only the nature of meaning; it is rather the role of the mind in human public affairs, linguistic as well as other; and there, as in semantics, I for one would not trust mindlessness.

Chapter 6

Conservative Rationalism II

6.1 The Old Sorcerer's Supervenience Chain

Reductive behaviourism, as we have seen, collapses to behavioural holism; we shall now turn to see that Fodor's computational mentalism, together with the referential theory of meaning and nomic theory of reference, collapses to reductive behaviourism, so that Fodor, willy nilly, ends up with the holists. In Section 1.2, I described Fodor's account of the mind in terms of a hypothetical supervenience chain that links organisms with their environment in this sense: identical organisms instantiate the same computational mental system; and identical mental systems carry the same semantic contents; and identical semantic contents determine the same extensions and truth-conditions. I then rehearsed the familiar Twin-Earth arguments to the conclusion that no theory of mind structured by the supervenience chain can be correct, since the physical and social environment itself contributes to the semantic identity of mental states. Thus one could vary the environment whilst holding a computational system the same; and this shows that the sameness of computational states is not sufficient for the sameness of extensions or truth-conditions; which in turn shows that mental states cannot be identified with the computational states of a physical system. In short, the Twin-Earth arguments uphold the principle of extension-meaning supervenience, but reject the supervenience of the mind on a computational neural system, and in general on any physical system whatever. Fodor's (1987) response was that the supervenience of mind on brain is compatible with the supervenience of extension on meaning, *relative to a context.* A context is a relevantly local universe of discourse, or domain of interpretation, with respect to which an organism's mental symbols have their meanings fixed by certain nomic relations of reference; and the requirement of relevant localness ensures that a domain including both Earth and Twin Earth is not a context. In contrast, Fodor's (1994) response was that the Twin-Earth scenario is *not nomologically possible*, and therefore cannot tell against the nomic theory reference and referential theory of mental content.

In Section 1.4, I drafted my claim against Fodor thus: the sameness of computational form is not sufficient for the sameness of meaning taken referentially, whether one assumes a context-relative referential theory of meaning (as in Fodor (1987)), or context-free referential theory of meaning (as in Fodor (1994)); more generally, mental states do not determine their extensions and truth-conditions: the principle of extension-meaning supervenience must be rejected as a principle of semantic and hence psychological individuation. Contrary to Putnam and Burge, and in agreement with Fodor, I upheld the principle of mind-brain supervenience, that the sameness of brain necessitates the sameness of mind. So the dice were cast as follows: Putnam and Burge stick to the rule of extension-meaning supervenience but quit mind-brain supervenience; Fodor holds onto extension-meaning and mind-brain supervenience; I accept mind-brain supervenience, so that the sameness of organism ensures the sameness of mind, but reject extension-meaning supervenience.

There is one aspect in regard of which Fodor's theory of mind has never changed; namely, that no *definitional* account of reference can be correct; *i.e.*, an account saying that some or most *prima-facie* lexical symbols are semantically complex, and these determine their extensions, or refer, by comprising *necessary and sufficient conditions* for the membership in the extensions. He was therefore compelled to seek a *nomic* and *dispositional* account of reference (1987; 1990a, b; 1994), or at worst, a *causal* and *historical* account (*e.g.*, 1981). The latter option having proved untenable, he tried to work out the former; but there, either he could assume that all *prima-facie* lexical symbols are semantically simple, unstructured and unlearned, in which case he would have a trouble in accounting for the *differing causal powers* of such *co-referential* pairs of beliefs as that the Morning Star is red and that the Evening Star is red; or he could assume that some (most) *prima-facie* lexical symbols are *syntactically* complex, structured, and learned, but their meaning still consists in *nomic reference*, in which case he would be able to say that the beliefs that the Morning Star is red and that the Evening Star is red, though *co-referential* and thus *synonymous*, differ nevertheless in their causal powers because of differences in their syntax. The former option was explored in (1987; 1990a, b), the latter in (1994). The point to emphasise is that, as concerns my claim against Fodor, it does not matter at all whether he assumes that most or all *prima-facie* lexical mental symbols are both syntactically and semantically simple, unstructured and unlearned, or that most or all such symbols are syntactically complex, structured, learned, yet still semantically simple; what matters is only that he assumes the referential theory of meaning and nomic theory of reference. In other words, the following refutation of Fodor's computational mentalism will rest solely on his semantics.

In section 1.3, I set out Fodor's nomic theory of reference. Fodor (1987; 1990a, b) put forth what we may regard as a *three-phase* account of reference, including the *physical*, the *psychophysical*, and the *psychological* phase. The physical phase links instances of the property a symbol refers to with instances of a psychophysical property by a physical law; the psychophysical phase links instances of the psychophysical property with tokenings of a psychophysical representation by a psychophysical law; the psychological phase links tokenings of the psychophysical representation with tokenings of the symbol by a psychological law. In (1994), Fodor added what we may regard as the *sociological phase* of reference: here the psychological phase is divided between experts and laity, so that tokenings of the psychophysical representation are mapped into tokenings of the symbol in point of a psychological law *only in the minds of experts*, with laity relying on the experts in order to align their tokenings of the symbol with tokenings of the psychophysical representation, and hence with instances of the property referred to.

It is, I suspect, rather obvious that such a nomic theory of reference and referential theory of meaning is incompatible with Fodor's computational account of mind, since it cannot sustain the basic mentalistic requirement that the *syntactic sameness* of tokens of mental symbols must be sufficient for their *semantic sameness*; without that requirement, there is nothing left of semantic mentalism, and no criteria left for the semantic identity of *public* symbols except their overt dispositional use. But if it is not yet obvious, this chapter will make it so.

I concluded Section 1.3 by pointing out a mistake in Fodor's (1987) account of error in terms of the alleged asymmetric dependence of misrepresentation on veridical representation. I argued, in summary, that Fodor tacitly presupposes in the account what he sets out to prove: namely, that tokenings of α which are nomologically dependent both on instances of $<\gamma>$ and on instances of $<\delta>$ do nevertheless refer to the property $<\gamma>$ rather than the property $<\delta>$ or $<\gamma \text{ or } \delta>$; which is to say, that Fodor's account of error by asymmetric dependence is circular.

In the next section, I will resume the argument just outlined. I will consider Fodor's (1990b) reply to a similar, though not quite the same, argument reportedly due to Ned Block. Then I will extract a non-circular account of asymmetric dependence one might plausibly attribute to Fodor (1990b). This will allow me to tie the problem of misrepresentation, and Fodor's solution to it, with the issue of accounting in greater detail for the *sociological phase* of reference. I will work out such an account in Section 6.3, and show that it follows from Fodor's computational mentalism, because of the sociological phase of reference, that mental symbols cannot determine their extensions and truth-conditions, in that syntactically identical symbol-tokens need not have identical extensions. Computational mentalism cannot therefore satisfy the elementary mentalistic requirement that syntactically identical symbol-tokens must be semantically identical; hence it collapses to behaviourism and, inevitably, to behavioural holism in respect of meaning.

6.2 Misrepresentation and Asymmetric Dependence

Here is another way of putting my argument in Section 1.3.2. Fodor sets up his asymmetric dependence solution to the disjunction problem for tokens of English rather than Mentalese:

> Consider the following situation: I see a cow which, stupidly, I misidentify. I take it, say, to be a horse. So taking it causes me to effect the tokening of a symbol; viz., I say 'horse.' Here we have all the ingredients of the disjunction problem (set up, as it happens, for a token of English rather than a token of Mentalese; but none of the following turns on that). So, on the one hand, we want it to be that my utterance of 'horse' means *horse* in virtue of the causal relation between (some) 'horse' tokenings and horses; and, on the other hand, we *don't* want it to be that my utterance of 'horse' means *cow* in virtue of the causal relation between (some) 'horse' tokenings and cows... [B]ut how are we to get what we want? (1987: 107)

There follows the asymmetric dependence answer:

> ... misidentifying a cow as a horse wouldn't have led me to say 'horse' *except that there was independently a semantic relation between 'horse' tokenings and horses...* Whereas, by contrast, since 'horse' does mean *horse*, the fact that horses cause me to say 'horse' does not depend upon there being a semantic — or, indeed, any — connection between 'horse' tokenings and cows.
> (*ibid.*: 107–108).

What interests us is Fodor's claim that nothing turns on setting up the account for tokens of English rather than Mentalese; for that is clearly false. We know of the English word "horse", as opposed to the hypothetical extension-determining mental symbol **horse**, that it refers to horses, or the species <horse> (*horse*, in Fodor's notation). We do not know only how "horse" succeeds in referring to <horse>. In contrast, supposing we could find a token of the semantically simple Mentalese symbol **horse** in the mind/brain, not only would we not know how it refers, we would also not know what it refers to — whether it refers to <horse> or <cow>, or <horse or cow>, or whatever — and therefore, on the referential theory of meaning, what it means. It follows that Fodor cannot allow himself, on pain of circularity, of the supposition that (the hypothetical extension-determining symbol) **horse** refers to <horse> rather than <cow> or <horse or cow>. So something does turn on Fodor's setting up the asymmetric dependence 'solution' to the disjunction problem for tokens of English rather than Mentalese; doing so hides Fodor's mistake.

Fodor (1990b: 111–113) alleges to have refuted a similar argument reportedly due to Ned Block. We shall now turn to what Fodor says is Block's argument, and to Fodor's reply to it. Block's argument is this:

> Look, your theory comes down to: "cow" means *cow* and not *cat* because, though there are nomologically possible worlds in which cows cause "cow"s but cats don't, there are no nomologically possible worlds in which cats cause "cow"s but cows don't. But such face plausibility as this idea may have depends on equivocating between two readings of "cow"... If you mean by "cow" something like *the phonological/orthographic sequence #c^o^w#*, then there's just no reason at all to believe the claim you're making. For example, there is surely a possible world in which cows don't cause *#c^o^w#*s but trees do, viz., *the world in which #c^o^w# means tree*. So, if when you write "cow" what you mean is *#c^o^w#*, then it clearly can't be nomologically necessary in order for "cow" to mean *cow* that nothing causes "cows" in worlds where cows don't... So, to put it in a nutshell, if you read "cow" orthographically/phonologically the claim that "cow" means *cow* because "cow"s are asymmetrically dependent on cows is false; and if you read "cow" morphemically the claim that "cow" means *cow* because "cow"s are asymmetrically dependent on cows is circular. Either way, it's a claim that seems to be in trouble. (Fodor 1990b: 111–112)

A couple of minor observations before turning to Fodor's reply:

(i) In referring, at the end of the quotation, to his own theory as 'the claim that "cow" means *cow* because "cow"s are asymmetrically dependent on cows', Fodor mixes up the terms "asymmetrically dependent" and "nomologically dependent"; see also Fodor (1990b: 113) for a similar mix-up. This, of course, is not a serious worry.

(ii) Block's argument is not quite the same as mine. Mine does not rest on the positing of a possible world in which **horse** 'means' *cow* (or **cow** 'means' *tree*, or whatever). Rather, my argument follows the set-up of the disjunction problem — according to which, plausibly, both cows and horses cause tokenings of **horse** — and shows that there is no asymmetric dependence, assuming the nomic theory of reference, between the nomological dependence of tokenings of **horse** on instances of <horse> and the nomological dependence of tokenings of **horse** on instances of <cow>; that is, the argument shows that, according to the nomic theory of reference and referential theory of meaning, the hypothetical extension-determining symbol **horse** must 'mean' <horse or cow>, not either <horse> or <cow>. In contrast, Block and Fodor argue on the basis of the possible-worlds construal of nomological dependence, having thus to decide what nomologically possible worlds there are, which are nearer to the actual world than others, and so forth. As we shall see anon, this may be partly responsible for Fodor's confusion.

In his reply to Block, Fodor argues that:

> Block is, of course, perfectly right that for the purposes of a naturalistic semantics the only nonquestion-begging reading of "cow" is *#c^o^w#*. Henceforth be it so read. However, the asymmetric dependence proposal is that *all else being equal*, breaking cow→"cow" breaks *X*→"cow" for all *X*. Correspondingly — to put the point intuitively — what’s wrong with Block’s argument is that all else *isn’t* equal in the worlds that he imagines. To get those worlds, you need to suppose *not only* that cow→"cow" is broken, *but also and independently* that tree→"cow" is in force. It’s this independent supposition that violates the ’all else equal’ clause. (1990b: 112-113)

Is the reply valid? Since the disjunction problem is now discussed by reference to cows and trees, rather than cows and horses, it does appear that Block’s supposition that ’tree→"cow" is in force’ is an independent supposition violating Fodor’s ’all else being equal’ clause. But the supposition is not independent. This becomes clear when we stick to the set-up of the disjunction problem, and consider the problem by reference to horses and cows rather than cows and trees. Our initial supposition for the disjunction problem was that, plausibly, both horses and cows cause tokenings of **horse**; and that supposition generated the disjunction problem. Hence it follows that when we suppose, in addition, that horse→"horse" is broken, the supposition that cow→"horse" is in force remains untouched. In other words, *mutatis mutandis*, Block’s supposition that tree→"cow" is in force is not independent, and does not violate Fodor’s ’all else being equal’ clause. So Fodor’s account of error is still circular and begs the question of semantic individuation. There is no asymmetric dependence, given the nomic theory of reference and referential theory of meaning in his computational mentalism, between veridical and unveridical representation. So far at least, Fodor gives us only an illusion of such a dependence, due to his setting up the asymmetric-dependence account for tokens of English rather than Mentalese.

However, and this is where the more interesting part of my case against Fodor begins, the asymmetric dependence account can be made (not, so I will argue, correct but) non-circular, by including a non-circular account of how the alleged asymmetric dependencies among the causal links between tokenings of mental symbols and instancings of environmental properties are generated. Moreover, Fodor (1990b) gives us some idea regarding the way such an account should go. To grasp the idea, we must begin with Fodor’s (1990b) slightly modified referential theory of meaning and nomic theory of reference, which is as follows.

> I claim that "*X*" means *X* if:
>
> 1. ’*X*s cause "*X*"s’ is a law.
> 2. Some "*X*"s are actually caused by *X*s.

> 3. For all Y not$=X$, if *Y*s qua *Y*s actually cause "*X*"s, then *Ys causing "X"s* is asymmetrically dependent on *Xs causing "X"s*. (1990b: 121)

Fodor's presentation here is rather telegraphic. The phrase '"*X*" means *X* if ...' is to be read as 'the meaning of the mental symbol "*X*" consists in its referring to the property *X*, and "*X*" refers to *X* if ...'. Likewise, the expression '"*X*"s' is to be read as 'tokens of the symbol "*X*"', and the expression '*X*s' is to be read as 'instantiations of the property *X*'. The rest is hopefully self-explanatory.

I will briefly comment on Fodor's clauses 1 and 2, to explain where the slight difference between his (1987) and (1990b) accounts of meaning is, and then I will concentrate at greater length on clause 3.

Re 1. Fodor claims that a theory of content such as that given by 1-3...

> has the desirable property of not assuming ... that there are circumstances — nomologically possible and naturalistically and otherwise nonquestion beggingly specifiable — in which it's semantically necessary that only cows cause "cows". Nor does it assume that there are nonquestion-beggingly specifiable circumstances in which it's semantically necessary that *all* cows would cause "cows". [Fodor's footnote:] Compare *Psychosemantics* (1987), in which I took it for granted — wrongly, as I now think — that an information-based semantics would have to specify such circumstances. As far as I can tell, I assumed this because I thought that any informational theory of content would have to amount to a more or less hedged version of 'all and only cows cause "cow"s'... (1990b: 91)

Fodor is certainly right in saying that his (1990b) theory does not assume that 'there are circumstances in which it's semantically necessary that *only* cows cause "cow"s'; and he might say the same about his (1987) theory, since in both cases the 'only'-problem is catered for by the asymmetric-dependence account of error. However, he is wrong in saying that, unlike the (1987) version, his (1990b) theory does not assume that 'there are circumstances in which it's semantically necessary that *all* cows would cause "cow"s'. For such circumstances are presupposed in clause 1 of the (1990b) theory: one cannot suppose that '*X*s cause "*X*"s' is a law, without assuming that under some circumstances all *X*s would cause "*X*"s. Fodor might perhaps reply that I misread his claim: his claim is that the (1990b) version does not assume there being 'circumstances in which it is *semantically necessary* that all cows would cause "cow"s'. But surely Fodor's term "semantically necessary" means "nomologically necessary", where "nomologically" pertains to the causal laws linking instances of environmental properties with tokenings of mental symbols. Hence again Fodor cannot escape the obligation to specify the circumstances wherein *all* instances of a property are causally sufficient for tokenings of the corresponding mental symbol.

Re 2. Fodor says that clause 2, that some "*X*"s are actually caused by *X*s, "invokes the actual history of "*X*" tokens as constitutive of the meaning of "*X*" and thereby ... rules out '"horse" means Twin-horse', '"water" means XYZ', and the like" (1990b: 121). He takes this to resolve the problems raised by the Twin-Earth examples. *Q.v.*: "Any property that *doesn't* actually cause "*X*"s ipso facto fails to meet condition 2; that's why "water" doesn't mean XYZ according to the present account" (*ibid.*: 122). This is a mistake. In assuming that instances of XYZ do not, as a matter of fact, cause tokenings of **water**, Fodor begs the question against Putnam. For Putnam's argument is that, for all we know, there might be a Twin Earth in outer space where the mental symbol **water** is causally occasioned by XYZ rather than H_2O. Fodor cannot simply assume without circularity that **water** is never actually caused by XYZ.

Re 3. Fodor claims that clause 3, the asymmetric-dependence account of error, "allows symbols to be *robust* with respect to their actual histories of tokening as well as with respect to their counterfactual histories. That is. it allows tokens of a symbol actually to be caused by things that are not its extension" (1990b: 122). In other words, mental symbols are *robust* because their unveridical tokenings are asymmetrically dependent on their veridical tokenings; and robustness consists in that tokenings of the symbols mean whatever they do despite being nomologically dependent on instances of properties other than those that define their extensions (as well as those that define their extensions). The question is, why should we believe that Fodor's hypothetical extension-determining mental symbols are robust in the requisite sense? Why should we believe that unveridical tokenings of such symbols are asymmetrically dependent on their veridical tokenings? Fodor's answer, or clue to an answer, is as follows. (I will ask the reader to bear with the unusual frequency of quotations in the ensuing few paragraphs; my sole excuse is that, without them, it is impossible to make clear just what Fodor's position is. Readers familiar with Fodor (1990b) will perhaps appreciate the point.)

To begin with, Fodor invites us to work with linguistic rather than mental symbols, as he says, "in order to keep the facts as much as possible out in the open" (1990b: 96). Then he offers us an example of an asymmetric dependence of the linguistic practice of paging people on the linguistic practice of naming people: "... there's the business of having people paged. How it works is: Someone calls out "John" and, if everything goes right, John comes. Why John?... Well, because the practice is that the guy who is to come is the guy whose name is the vocable that is called" (*ibid.*: 96). Fodor next conjectures that all linguistic practices depend asymmetrically on the practice of naming: "Some of our linguistic practices presuppose some of our others, and it's plausible that practices of *applying* terms (names to their bearers, predicates to things in their extensions) are at the bottom of the pile" (*ibid.*: 97). So far, none of this tells us how the alleged asymmetric dependencies are generated, and why linguistic, not to mention psychological representations are, as Fodor says, robust. He then goes on to consider the following problem: "...we've been construing robustness by appeal to asymmetric dependencies among *linguistic practices*. And linguistic *practices* depend on linguistic *policies*... Since, however, being in pursuit of a policy is being in an intentional state, how could asymmetric dependence among linguistic practices help with the naturalization problem?" (*ibid.*: 98). His answer is:

> Perhaps the policies per se aren't what matters for semantics; maybe all that matters is the patterns of causal dependencies that the pursuit of the policies give rise to... The point is, if the asymmetric dependence story about robustness can be told just in terms of symbol-world causal relations, then we can tell it *even in a context where the project is naturalization*. No doubt, it's the linguistic policies of speakers that give rise to the asymmetric causal dependencies in terms of which the conditions for robustness are defined; but the conditions for robustness *quantify over* the mediating mechanisms, and so can be stated without referring to the policies; hence their compatibility with naturalism. (1990b: 99-100)

What interests us in the foregoing is not so much that 'being in pursuit of a policy is being in an intentional state', and whatever problems this might raise for naturalistic theory of mental content, but rather that 'it's the linguistic *policies* of speakers that give rise to the asymmetric causal dependencies in terms of which the conditions for robustness are [to be] defined'. For, in other words, Fodor's position is that the asymmetric dependencies of the various linguistic practices upon the fundamental linguistic practice of applying terms to their extensions are generated by the speakers' *being in pursuit of linguistic policies*, and that being in pursuit of a linguistic policy is a certain complex *mental state*. We now have some non-circular explanation — accepting as we may Fodor's contention that 'the policies per se aren't what matters for semantics' — of asymmetric dependence, albeit an explanation couched in terms of linguistic rather than psychological symbols. What then of asymmetric dependence concerning mental symbols? Here is what Fodor has to say:

> [I]f it's the causal patterns themselves that count, rather than the mechanisms whose operations give rise to them, then perhaps our *mental* representations can be robust just in virtue of asymmetric dependencies among the causal patterns that our concepts enter into. That is, perhaps there could be mechanisms which sustain asymmetric dependencies among the relations between mental representations and the world, even though, patently, we have no policies with respect to the tokenings of our mental representations. If that were so, then the conditions for the robustness of

> linguistic expressions and the conditions for the robustness of mental representations might be *identical* even though, of course, the mechanisms in virtue of whose operations the two sorts of symbols satisfy the conditions for robustness would be very, very different. (1990b: 100)

This is as much as Fodor says concerning the nature of robustness and asymmetric dependence. Let me now elaborate a little on the account. If, contrary to my argument in Section 1.3.2, there is a non-circular account of asymmetric dependence for mental symbols, then it must rely on the acquisition of certain complex mental states, or 'policies' of a sort, though Fodor says that 'patently, we have no policies with respect to the tokenings of our mental representations'. (Why this is patent is not clear; on the contrary, there are plenty of policies as to what people should think and when.) These complex mental states can be regarded as *internalised theories* which function to map, in the *psychological phase* of reference, certain sensory or other evidence into tokenings of the symbols. For example, if there is a non-circular account of the asymmetric dependence of cow-caused tokenings of the (hypothetical extension-determining) symbol **horse** upon horse-caused tokenings of **horse**, then it must rely on the acquisition of an internalised theory about horses which works to map, in the psychological phase of the causal link between horses and tokenings of **horse**, the sensory or other evidence one gets of horses into tokenings of **horse**. Once such a theory or internal 'policy' has been built, cow-caused **horse**-tokenings indeed may asymmetrically depend on horse-caused **horse**-tokenings.

If this is Fodor's intended account of asymmetric dependence, then his position is not circular; moreover, the account accords well with his four-phase nomic theory of reference and in general with his computational mentalism. Suppose then that there are such internalised theories that cater for the psychological phase of reference (at least for experts, if not laity), and that generate the asymmetric dependencies of unveridical upon veridical tokenings of mental symbols. Still, so I will argue, the referential theory of meaning and the nomic theory of reference let Fodor down on the disjunction problem, and make his computational mentalism collapse to behaviourism and behavioural holism.

6.3 The Sociological Phase of Reference

The chief source of trouble for Fodor's computational mentalism and the associated nomic theory of reference is that most of the nomic links connecting instantiations of aspects of the mind's environment with tokenings of mental symbols cannot but run *via* what I have called "the sociological phase" of reference. The sociological phase can be regarded as a *learning* phase; it is the phase in which laity learn from experts theories about such properties as being a proton, elm, gold, and so on. The nomic theory of reference, to repeat, says that a mental symbol refers to a property (or the extension defined by the property) if there is a canonical causal chain that links instances of the property with instances of a psychophysical property by a physical law; that links the latter with tokenings of a psychophysical representation, in an expert's mind, by a psychophysical law; and that links, in the expert's mind, tokenings of the psychophysical representation with tokenings of the symbol by a psychological law, *via* an internalised theory the expert knows about the property. In the sociological phase, the canonical causal chain links tokenings of the symbol in the expert's mind with tokenings of the *type-identical* symbol in *lay minds*; that is, in the mind of *every* lay speaker who can be said to have acquired the symbol and know the meaning of the corresponding word. There are, in principle, only two ways in which the sociological phase of the nomic chain could occur: either the lay speakers *learn* from the expert an *internal theory* about the property referred to by the symbol, that is *sufficient* to map whatever evidence they may gather of instances of the property into their own tokenings of the symbol; or else the internal theory they learn about, say, protons is that protons are whatever the experts on whom they rely call "proton". We shall look into these alternatives anon.

The notion of the sociological phase as a learning phase of reference comports with Fodor's computational mentalism and his referential theory of content. According to Fodor, the mind is a system of operations on the symbols of a Mentalese language; the semantically simple symbols are innate and universal for the species; and the meaning of a simple symbol consists in the nomic dependence of its tokenings on instances of its extension. Furthermore, the notion comports well with Fodor's account of error in terms of asymmetric dependence of misrepresentation on veridical representation, as well as with his account of the origin of the alleged asymmetric dependencies in terms of *internalised theories*, or 'internal policies'.

Given so much, consider once again the elm-beech example due to Putnam (1975): suppose Jane's representation of elms is the same as her representation of beeches, namely, **common deciduous tree**; the representations are identical, yet the extensions of the natural kinds <elm> and <beech> still are, respectively, the set of elms and the set of beeches. We noted in Section 1.2.2 that this argument does not refute Fodor's position since it rests on the supposition that lexical mental symbols are semantically complex descriptions that would determine their extensions only if they comprised necessary and sufficient conditions for the membership in the extensions; and that supposition need not hold, since a *nomic* rather than *definitional* or description-based account of reference for such symbols may in fact be true.

Let us now contemplate a modified version of the elm-beech case, adapted to suit the sociological phase of the nomic theory of reference. The nomic theory has it that the simple symbol **elm** refers to the property <elm>, or extension {elm}, because tokenings of **elm** are nomologically linked to instancings of <elm>. The physical phase of the link maps an instancing of <elm> in the environment into the corresponding instancing of a psychophysical property, say, the 'look' of the particular elm. The psychophysical phase maps the latter into a tokening of a psychophysical representation, in the sensorium of an expert observer under psychophysically good circumstances. The observer, being one of the experts, a dendrologist, knows a theory that maps the tokening of the psychophysical representation into a tokening of the simple symbol **elm**. So far, so good. A similar account holds, *mutatis mutandis*, for the nomic link between instancings of <beech> and tokenings of **beech** in experts' minds.

In the sociological phase of the link between <elm> and **elm**, lay speakers learn from experts a theory about elms which maps, in their psychological phase, whatever evidence they gather of elms into their tokenings of **elm**; likewise for the sociological phase of the link between <beech> and **beech**. Grant now that what laity learn from experts about elms and beeches falls short of expert knowledge; that, in other words, there is a division of *epistemic* labour (though not a division of semantic labour); and suppose that the theory most lay speakers learn from experts about elms and beeches is more or less the same: *viz.*, **common deciduous tree**. But then, there is no reason why in laity's minds **elm** should not refer to <beech>, or **beech** to <elm>, or both **elm** and **beech** to <elm or beech>. For since the internalised theories about elms and beeches in laity's minds are much the same, the psychological phase of the nomic link between instancings of <elm> (or <beech>) and tokenings of **elm** (or **beech**) will map whatever evidence laity gather of either elms or beeches into tokenings of **elm**, as well as into tokenings of **beech**.

Fodor might object that <beech>-caused tokenings of **elm** will turn out asymmetrically dependent on <elm>-caused tokenings of **elm** (and analogously for tokenings of **beech**). He might argue thus: 'your assumption that the theory most lay speakers acquire from experts about elms and beeches is the same — *viz.*, **common deciduous tree** — need not and should not be granted. More plausibly, the theory laity learn from specialists about elms might be **common deciduous tree which experts on whom I rely call "elm"**, and the theory they learn about beeches might be **common deciduous tree which experts on whom I rely call "beech"**. If so, there is a reason why, in laity's minds, **elm** should not refer to <beech>, or **beech** to <elm>, or both **elm** and **beech** to <elm or beech>. Since the theories about elms and beeches are not the same, whenever the psychological phase between instancings of <elm> (or <beech>) and tokenings of **elm** (or **beech**) will map, unveridically, the evidence laity gather of elms into tokenings of **beech** (or the evidence of beeches into tokenings of **elm**), the unveridical causal links will be asymmetrically dependent on the veridical links: for, given our assumptions, there could be no <elm>-caused **beech**-tokenings without there being <beech>-caused **beech**-tokenings, though there could be <beech>-caused **beech**-tokenings without there being any <elm>-caused **beech**-tokenings; and similarly, *mutatis mutandis*, for <elm>-caused and <beech>-caused **elm**-tokenings. So the nomic theory of reference and referential theory of mental content still stand!'

The foregoing is a *fodorised* version of an argument which Putnam (1988: 26) attributes to Searle (1983); see Searle (1983: 201–202) for the argument Putnam presumably has in mind. I say "fodorised" since the argument Putnam credits to Searle rests on the supposition that such mental descriptions as **common deciduous tree which experts on whom I rely call "elm"** just are laity's mental representations of elms. By contrast, the fodorised version has it that these descriptions are *internalised theories* mapping instances of <elm> into tokenings of the *semantically simple* symbol **elm** in the psychological phase of reference of lay speakers, and generating the asymmetric dependencies of unveridical on veridical tokenings of **elm**.

Does the objection save Fodor's position? The gist of the objection is this: the mental description or theory laity acquire from experts about elms is not the same as the description they acquire about beeches, because — though all they seem to know about elms and beeches is much the same (*viz.*, that elms and beeches are common deciduous trees) — they also know that elms are not beeches and beeches are not elms, and that elms are called "elm" and beeches "beech" by the experts on whom they rely; and this additional knowledge is what generates the asymmetric dependencies of unveridical on veridical tokenings of **elm** and **beech**.

But can the additional knowledge, that elms and beeches are not the same, and that elms are called "elm" whereas beeches "beech", generate the asymmetric dependencies Fodor's account needs? Clearly, it can not. Regarding the knowledge that elms are not beeches and beeches are not elms, if it is a part of the mental description which laity learn about elms, then surely it is also a part of their description of beeches (and conversely). So far, there is no reason to believe that the descriptions differ. Consider now the knowledge that elms are called "elm" whereas beeches "beech". Can this knowledge generate the needed asymmetric dependence of <beech>-caused tokenings of **elm** upon <elm>-caused tokenings of **elm** (and of <elm>-caused tokenings of **beech** upon <beech>-caused tokenings of **beech**)?

We may express the question a little more explicitly. The description which laity learn from experts about elms is **common deciduous tree which experts on whom I rely call "elm"**; can the phrase **which experts on whom I rely call "elm"** generate an asymmetric dependence of <beech>-caused upon <elm>-caused tokenings of **elm**? At first glance, it may seem the answer is "yes". For, to put it in Fodor's nomenclature, experts together with laity pursue a certain 'linguistic policy': the policy of calling elms "elm". Further, their being in pursuit of the policy amounts to, at least in laity's case, their having such phrases as **which experts on whom I rely call "elm"** embedded in their mental descriptions of elms, the descriptions which map, in the psychological phase of reference, instancings of <elm> into tokenings of **elm**. Hence it seems to follow that there could not be <beech>-caused **elm**-tokenings unless there were <elm>-caused **elm**-tokenings, whilst there could be <elm>-caused **elm**-tokenings without there being any <beech>-caused **elm**-tokenings; in other words, that <beech>-caused tokenings of **elm** (generally, unveridical tokenings of **elm**) are asymmetrically dependent on the veridical <elm>-caused tokenings of **elm**.

At second glance, though, the answer to our question is "no". For there is nothing in the phrase **which experts on whom I rely call "elm"** to ensure that the psychological phase will map instances of <elm> into tokenings of **elm** rather than, say, **beech** *in laity's minds*. (The same, of course, does not hold for experts' minds, who know a theory about elms which is sufficient to map instancings of <elm> into tokenings of **elm**.) In fact, the phrase **which experts on whom I rely call "elm"** cannot assist at all to ensure that the basic mental description — *viz.*, **common deciduous tree** — will not map instancings of <elm> into tokenings of **beech**, **maple**, or any other representation of a common deciduous tree. Even if there were no <beech>-caused or <maple>-caused **elm**-tokenings (*e.g.*, if there were no beeches or maples around), still, the psychological phase — deploying as we suppose the description **common deciduous tree which experts on whom I rely call "elm"** — could map the evidence laity gather of actual elms into tokenings of **beech**, **maple**, *etc.*; there is no psychological law to ensure that the phrase **which experts on whom I rely call "elm"**, settled in accord with public linguistic convention, will discriminate between tokenings of **elm**, **beech**, or **maple**, even with respect to identical or similar psychophysical evidence laity may gather of actual elms. It follows, assuming Fodor's nomic theory of reference and referential theory of meaning, that the disjunction problem reappears: there is no reason why, in laity's minds, **elm** should not refer to <beech>, or **beech** to <elm>, or both **elm** and **beech** to <elm or beech>.

The disjunction problem is a symptom of a general breakdown of the mentalistic principle of meaning-syntax supervenience, requiring that syntactically identical tokens of mental symbols be semantically identical. Failing this principle, Fodor's computational theory of mind, with its referential theory of meaning, can no longer work as a mentalistic semantical theory, either for its own hypothetical extension-determining mental symbols, or for public language; it can no longer claim that the meanings of public words, and generally the meaning-bearing states of human or other organisms, ultimately reduce to the meanings of mental symbols, and that public symbols derive their significance from mental symbols. Insofar as it still includes the referential theory of meaning and the nomic-dispositional theory of reference, these can now apply only to public symbols; but this brings Fodor's computationalism, as regards meaning, down to reductive behaviourism, and hence to behavioural holism.

We shall, in conclusion, look into some of Fodor's passing remarks on the import of the sociological phase of reference for his account of mind:

> Someone who is an Intentional Realist but not a behaviorist could ... embrace a Skinnerian semantics *for thoughts* while entirely rejecting Skinner's account of language. Here's how the revised story might go: There is a mental state — of entertaining the concept DOG, say — of which the intentional object is the property *dog*. (Or, as I shall sometimes say for brevity, there is a mental state that *expresses* the property *dog*). The fact that this state expresses this property reduces to the fact that tokenings of the state are, in the relevant sense, discriminated responses to instances of the property; i.e., instancings of the state covary with (they are 'under the control of') instancings of the property, and this covariation is lawful, hence counterfactual supporting.
>
> This account isn't behavioristic since it's unabashed about the postulation of intentional mental states. And it isn't learning-theoretic since it doesn't care about the ontogeny of the covariance in terms of which the semantic relation between dog-thoughts and dogs is explicated...
>
> The basic idea of Skinnerian semantics is that *all* that matters for meaning is "functional" relations (relations of nomic covariance) between symbols and their denotations [referents]. In particular, it doesn't matter *how that covariation is mediated*; it doesn't matter what mechanisms (neurological, intentional, spiritual, psychological, or whatever) sustain the covariation. This makes Skinnerian semantics atomistic in a way that Quineian semantics, for instance, isn't. (Fodor 1990a: 55–56)

The reader can see that Fodor regards his own account of meaning as, in a sense, Skinnerian. Skinner's semantic behaviourism differs from Quine's holistic version of it only in that it is a *reductive* behaviourism and thus 'atomistic'. So Fodor's theory of meaning and mind is, by his own acknowledgment, reductive behaviourism made mentalistic. Semantic mentalism, however, requires more than setting up reductive behaviourism for mental symbols. Fodor's mistake is that he never bothered to sort out just what classical semantic mentalism involves, as regards such issues as identifying the range of semantically simple symbols, distinguishing between empirical (*a posteriori*) and non-empirical (*a priori*) symbols, between the representation of nominal and of real or noumenal properties, and other issues (which we shall come to in Chapters 7–10).

This, though, is not the main point I am concerned with. More interesting is Fodor's claim that his Skinnerian semantics 'isn't learning-theoretic since it doesn't care about the ontogeny of the covariance in terms of which semantic relations are explicated'. There is something right, something wrong with this claim. The *referential theory of meaning* is indeed not learning-theoretic, since all that matters to it is *that* relations of reference obtain, not *how* they are mediated. But the *nomic theory of reference* is, as Fodor himself admits, learning-theoretic, because of the sociological phase of reference; yet he takes it for granted that this will have no effect on the *referential theory of meaning*:

> By the way, not just one's own skills, theories, and instruments, but also those of experts one relies on, may effect coordinations between, as it might be, "elms" in the head and elms in the field. That would be quite compatible with the meaning relation being both atomistic *and individualistic*, assuming, once again, the Skinnerian view that the conditions for meaning are purely functional and that they quantify over the mechanisms that sustain the semantically significant functional relations. Putnam (1988) argues that since appeals to experts mediate the coordination of one's tokens of "elm" with instances of *elm*, it follows that "reference is a social phenomenon." Prima facie, this seems about as sensible as arguing that since one uses telescopes to coordinate one's tokens of "star" with instances of *star*, it follows that reference is an optical phenomenon. That Putnam, of all people, should make this mistake is heavy with irony. For, it is Putnam who is always — and rightly — reminding us that "... 'meanings' are preserved under the usual procedures of belief fixation..."... I take this to be a formulation of anti-instrumentalist doctrine: the ways we have of telling when our concepts apply are *not*, in general, germane to their semantics. Why, I wonder, does Putnam make an exception in the case where our way of telling involves exploiting experts? (1990a: 83; note 6)

Putnam's answer to Fodor's question is, as we have seen in Chapter 2, that there is a division of *linguistic* or *semantic* labour among experts and laity; and it is a wrong answer. Nonetheless, there is a grain of truth in it which Fodor overlooks. The truth is that, because the *nomic theory of reference* requires an account of the sociological phase, and because the sociological phase involves a division of *epistemic* labour among experts and laity, and because the theories laity learn from experts about, for instance, elms and beeches are likely to be similar or identical (except for such functionally ineffective pieces of knowledge as that elms are called "elm" whereas beeches "beech"), and because, consequently, a version of the disjunction problem reappears for the nomic theory of reference, it follows that the *referential theory of meaning*, and with it Fodor's computational mentalism, breaks down.

It also follows that, in general, mental symbols do not determine their extensions and truth-conditions. For each symbol is either *semantically simple* or *semantically complex*; if most lexical symbols are complex, then the arguments from undefinability, such as the elm-beech case, apply; if most are simple (and the determination of extension is supposed to be nomic rather than definitional), then the argument from the sociological phase of reference applies. Either way, the hypothesis that symbols by their meaning determine their extensions is bound to fail. The Conservative principle that a difference in extension implies a difference in meaning, and the sameness of meaning necessitates the sameness of extension, must be rejected.

This outcome is not surprising, for the principle of extension-meaning supervenience — especially considered as a part of Fodor's supervenience chain, linking organisms *via* their mental contents to their environments — makes implicitly quite an extraordinary proposition: that organisms could, by thinking alone, find out how the environment is (and perhaps even, with a little thaumaturgy, manipulate the environment by thinking alone). This extraordinary proposition could be correct to some extent: classical mentalistic metaphysics of Descartes, and also some metaphysical aspects of Locke's position (see (1975: IV, X)), hold that certain fundamental features of what there is are discoverable (though certainly not manipulable) by thought alone. But this metaphysical mode of knowledge cannot extend to empirical and thus contingent propositions. Accordingly, neither Descartes nor Locke endorse the principle of extension-meaning supervenience. For them, most mental symbols or ideas represent *nominal* rather than *real* properties of the environment, and whether and to what extent some symbols may represent real properties is one of the deepest questions to ponder. Fodor's computational mentalism would have been to them an absurd doctrine.

Nor is rejecting the principle of extension-meaning supervenience any loss to classical term-based individuation of meaning, whether in the mental or public code. True, complex ideas may vary from person to person and, for each person, from time to time; but, as regards the mental code, this is simply to say that the composition of complex ideas is not fixed once and for all, and that complex ideas need not comprise necessary and sufficient conditions for the membership in their extensions; as regards the public code, it is to say that the meaning of a speaker's token of a word is fixed by the meaning of the idea the speaker uses the word to express *on that occasion*, and that communication depends on finding a common acceptation for each word within a linguistic community. Term-based semantics is part and parcel of mentalism; without mentalism, one is left with overt behavioural use of linguistic tokens, and there — finding reductive behaviourism not viable — one is doomed to behavioural holism.

Yet Fodor's supervenience chain could not help term-based mentalism a bit; and just as well, for it is not needed: we may consign it where it belongs, to the world museum of occult sciences. (But beware! Though his magic chain be broken, the sorcerer himself is still at large, and is reported to be hiding in a nearby possible world with an asymmetric relation of accessibility, which he hopes one day to reverse solely by an act of *de re* meaning!)

Chapter 7

The Classical Theory of Mind I

7.1 Five Ways of Defining Logical Modality

The Classical Theory of Mind stands and falls with a certain account of propositional logical modality: namely, the account in terms of clear and distinct ideas, which we shall elaborate in detail in this chapter. In modern Analytic Philosophy, there have been roughly four kinds of account of logical modality on offer. Firstly, the best known of these is the account in terms of *possible worlds*, according to which a proposition is, say, logically true just in case it is true in all possible worlds. Secondly, logicians have often favoured the account in terms of *deducibility* in a formal system, according to which a proposition is logically true just in case it is deducible as a theorem in such-and-such an axiomatic or deductive system. Thirdly, linguistically orientated philosophers have tended to think of logical or necessary truth as a matter of mere *linguistic*-cum-*behavioural convention* and resistance to revision. Fourthly, very much out of favour has been the notion that a proposition is analytically or logically true just in case it is *true by virtue of its meaning*. The first three accounts have been characteristic of Analytic Philosophy, and have been often inter-mingled: not uncommonly a philosopher and logician would hold that necessary truth is both a matter of truth in all possible worlds and of deducibility, and that it is all but a behavioural convention. The fourth account has been typically seen as the Conservative mentalistic position, and as irrevocably defeated by Quine's (1951) attack on the analytic-synthetic distinction. The CTM approach we are about to embark on, in terms of clear and distinct ideas, could be regarded as spelling out the allegedly discredited notion of analytic truth as truth by virtue of meaning. But I am not happy to regard it so; not because I would again have to combat Quine (for that would not be too difficult), but because the notion of analyticity as truth by virtue of meaning is so truncated as to be a foregone conclusion. In general, Quine's notions of classical mentalism — that the meaning of a word is an idea in the mind, that an analytic truth is truth by virtue of meaning, a synthetic truth is truth by virtue of fact, that 'the *a priori*' is knowledge independent of experience, 'the *a posteriori*' knowledge dependent on experience, and so forth — are left in so poor a state it is impossible to make any coherent sense of them.

Rather than being embroiled in such notions, I will turn to the Classical Theory of Mind itself; and, in order to make its exposition as plain as possible, I will design in this chapter a simple formal model of CTM, in which the mentalistic notions of meaning and logical modality can be clearly investigated. The following quotation will serve to set the theme:

> Another Faculty, we may take notice of in our Minds, is that of *Discerning* and distinguishing between the several *Ideas* it has... On this faculty of Distinguishing one thing from another, depends the *evidence and certainty* of several, even very general Propositions, which have passed for innate Truths; because Men over-looking the true cause, why those Propositions find universal assent, impute it wholly to native uniform Impressions; whereas it in truth *depends upon this clear discerning Faculty* of the Mind, whereby it perceives two Ideas to be the same, or different. (Locke 1975: II, XI, 1)

The propositions Locke speaks of are classical logically necessary truths: for example, that everything is identical to itself; that it is impossible for the same thing, at the same time, to be so-and-so and not to be so-and-so; and so forth. Although Locke's main purpose is to show that there is no need to suppose such propositions innate and universal for the human mind, his arguments for this claim reach beyond the concerns of psychology. Implicit throughout the *Essay* is a *logical theory* which Locke uses, on the one hand, to defend his psychology against the doctrine of innate propositional knowledge; and, on the other hand, to account for such properties of propositions as necessary — or, to put it in Locke's nomenclature — *demonstrative* truth.

The main goal of this chapter is to clarify what that implicit logical theory is, and how it can enlighten contemporary issues regarding the nature of logical modality, including the modality of conditional propositions and the validity of argument. However, my exposition of the theory will not strictly follow Locke's *Essay*. Section 7.2 will outline the basic tenets of the *Classical Theory of Mind* (CTM); and CTM will be taken to involve some common assumptions in the psychology of Descartes and Kant, as well as Locke. Section 7.3 will set out an account of *representation*, or the meaning of mental symbols, drawn primarily from Locke but common to CTM. Section 7.4 will review the classical theory of *epistemic evaluation*, or determination of truth-value, of propositions; specifically, those modes of evaluation whereby the mind acquires what Locke calls "intuitive" and "demonstrative knowledge". Section 7.5 will put forth an account, proceeding from the notion of intuitive and demonstrative knowledge, of such *modal properties* of propositions as logical necessity, contingency, possibility, *etc*. Section 7.6 will extend this account to cover the modalities of *conditional propositions*, and Section 7.7 will make use of the foregoing notions to form a CTM-based account of *deductive validity*. Section 7.8 will summarise and conclude the chapter with some reflections on the significance of these issues for the project of Analytic Philosophy, and on the role of Analytic Philosophy in the context of the classical tradition.

7.2 The Classical Theory of Mind

CTM rests on a certain construal of common-sense, introspective psychology: it regards the mind as a system of (instances of) *operations*, to be identified with believing, desiring, expecting, supposing, and so forth, on (tokens of) the sentences of a representational mental code, or *propositions*; and it regards propositions as composed of (tokens of) mental terms, or *ideas*. The following clauses, 7.2.1–7.2.5, will spell out — in as much detail as is necessary for our purposes — the key assumptions underlying CTM.

7.2.1 There is a *finite* basis of semantically *simple ideas* (taken as *types* of symbol), or simple terms of the mind's representational code.

Comments. Locke recognises several categories of simple ideas: there are ideas of sensation and ideas of reflection; in each class, there are ideas representing primary qualities, and those representing secondary qualities (there are no simple ideas of tertiary qualities); ideas of sensation are further classified as visual, auditory, olfactory, ideas of bodily states such as pain and hunger, and so on. I need not take a stance on the exact composition and size of the basis of simple ideas, but I will assume, in accord with not only Locke but also Descartes and Kant, that:

(i) each *basic* idea — *i.e.*, one belonging to the basis — represents a *mode*; that is, a non-relational, monadic property (so that there are no basic ideas standing for either relations or particular objects);

(ii) the basis is *empirical*, in that it comprises only ideas representing either sensory or reflective modes.

Formal model. In order to begin to construct a rudimentary formal system illustrating the logic of CTM, I will assume that the empirical basis consists of twenty simple ideas, signified by the first twenty letters of the alphabet:

A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T.

Further, I will assume that a basic idea may be tokened either as a grammatical *predicate*, or as a *subject*, or *object* (*i.e.*, grammatical object); and I will allow of an unlimited number of tokens, as opposed to types, of any idea (basic or other). When a basic idea is tokened as a subject or object (for example, in the proposition **blue is not identical to yellow**, the basic ideas **blue** and **yellow** are tokened, respectively, as subject and object), I will represent it by one of the twenty lower-case letters: a, b, c, ... When it is tokened as a predicate, I will represent it by an upper-case letter followed by a variable: Ax, Ky, Qz, ... The last three letters of the alphabet, with numerical subscripts if necessary, will be reserved as variables: x, y, z, x_1, y_1, ... (The remaining letters — u, v, w, with subscripts if needed — will be used as individual constants; see Section 7.5.1.)

Finally, I will suppose that the mind has an *input analyser*, the function of which is to distribute variables over predicate tokens in an appropriate way: *e.g.*, when an organism perceives, at the same time, an apple which is red and green, and a pear which is brown and yellow, its mind will token the ideas **red** and **green** (say) x-wise, and the ideas **brown** and **yellow** (say) y-wise. In this way, the input analyser will *itemise* or, if you please, *articulate* the environment which the mind thinks about.

7.2.2 There is a *generative mechanism* comprising operations for the production of infinitely many *complex ideas* from the empirical basis.

Comments. Locke distinguishes three generative operations for complex ideas: namely, composition, abstraction, and comparison. Composition works to combine several ideas into one compound idea; abstraction works to separate an idea from one or more complex ideas that contain it in common; and comparison works to generate ideas of relations — for example, the idea of identity — by comparing two or more other ideas in regard of their semantic content, or meaning. In Locke's view, these operations are not *idea-laden*: that is, though they are operations *on* empirical ideas, they do not involve or require any further, *non-empirical* ideas in order to function.

Formal model. For the sake of my model of CTM, I will depart from Locke in the following two respects.

Firstly, contrary to his *empiricism* — according to which there are no simple ideas except those comprising the basis — I will assume that the generative mechanism is idea-laden, consisting of two simple, non-empirical ideas representing the standard truth-functions: *viz*., $\wedge$ and $\neg$; and that other ideas of truth-functions — $\vee$, $\supset$, $\equiv$, *etc*. — are definable from the two simples. The mechanism will then generate complex ideas by variously compounding the basic empirical predicate-tokens, much as do the formation rules of predicate logic:

$(Ax \wedge Bx)$, $(Fy \supset \neg Gz)$, $((Kx \vee Jy) \equiv (\neg Jy \supset \neg\neg Kx))$, and so on.

Secondly, I will suppose that the Lockian operation of *comparison* works to generate the idea of identity, $=$, by comparing any two tokens of the same type of idea; and I will regard $=$ as a simple idea, though not as belonging to the empirical basis of simple ideas. In other words, I will suppose — contrary to Locke's empiricism, which takes all ideas derived by comparison, including the idea of identity, as complex — that the mind has the capacity for a hierarchy of simple ideas, beginning with the basic simple ideas, and forming further, non-empirical, relational simple ideas, among them the idea of identity, by the operation of comparison. (It is plausible to think of comparison as forming complex as well as simple relational ideas, but we need not include such ideas in the model.)

7.2.3 There is a generative mechanism comprising operations for the production of infinitely many *propositions* from the stock of simple and complex ideas.

Comments. These operations organise ideas into propositions, and are necessary for the mind's capacity to represent the external or internal environment as instantiating states of affairs, to judge the environment to be so-and-so, *etc*.

Formal model. I will take it that the generative mechanism for propositions is, like that for complex ideas, *idea-laden*, comprising the non-empirical, simple idea Σ, expressible by the word "some"; and I will allow three kinds of proposition to be formed:

(i) *propositions of identity* between any two basic ideas: *e.g.*, $a=a$, $a=b$, $j=k$, ...

(ii) *quantified propositions*: $(\Sigma\delta)\Gamma\delta$, with δ a variable, and $\Gamma\delta$ any predicate-token, simple or complex, in which all occurrences of δ are free, and all occurrences of other variables, if there are any, bound.

(iii) *compound propositions*: $\neg\alpha$, $(\alpha * \beta)$, where α and β are any propositions, and $*$ is any dyadic truth-functional idea.

$(\mathrm{A}\delta)\Gamma\delta$ — where A (Greek *Alpha*) is the idea of universal quantification, expressible by "all" — will be defined as usual by $\neg(\Sigma\delta)\neg\Gamma\delta$. Clause *(ii)* should be read as allowing that $\Gamma\delta$ may be a complex predicate-token involving other quantifiers: *e.g.*, $(Fx \wedge (\mathrm{A}y)(Fy \supset y=x))$, *etc*. The provisions for $\Gamma\delta$ are to ensure that there are no propositions with vacuous quantifiers, or quantifiers binding no variable in their scope, and no propositions with free variables.

The idea Σ could be represented by the standard Russellian sign "$\exists$", and A by "$\forall$"; but there are good reasons for changing the notation. "$\exists$" is typically regarded as an *existential* quantifier, as well as an *indefinite* quantifier expressible by "some"; so, if I retained Russell's notation in my model of the mental code, I would have to take "$\exists$" for an idea of existence, expressible by "is" or "exists"; however, the notion SOME is very different from the notion IS or EXISTS, and this should be reflected in the ideas one introduces into the formal model: one should have a distinct idea for the notion SOME, and a distinct idea for the notion IS. In the present *introductory* model, I will use Σ for SOME and A for ALL, and I will allow the notion IS to be merely *implicit* in the model, without being explicitly expressed by a distinct idea (see Section 7.3.3 for the semantics of Σ and A); but in a fuller model, the idea of existence should be explicitly distinguished from Σ. (A further reason for changing the Russellian notation is that a part of this project is to restore logic upon its psychological foundations; in other words, to rescue it from the Russellian anti-mentalistic, upside-down and backward stance; and what better measure to begin with, than to turn the symbols face-to-front and on their feet again.)

Both the generative operations for complex ideas and those for propositions comprise ideas, simple and complex, which are *non-empirical*, and which become psychologically active — in the formation of complex empirical ideas and propositions — only when some basic ideas are occasioned by, and formed in the mind in response to, experiential stimulations. In this sense, I will regard these operations as constituting, in part, the *a priori* faculty of the mind; and the non-empirical ideas laden in the operations — together with the semantically simple, non-empirical idea of identity — I will regard as *a priori* ideas. Thus I will have two kinds of idea involved in my model of CTM:

(a) *empirical* (or *a posteriori*) ideas, which are either of sensation or of reflection, and either simple or constructed from the simples;

(b) *a priori* ideas, which are non-empirical, either simple or complex, and the function of which is, in part, to organise simple empirical ideas into complex ideas and propositions.

In the following sections, I will gradually specify the *a priori* faculty further. In Section 7.5.1, I will treat of the mind's *a priori analytic* knowledge in the formal model; *a priori synthetic* knowledge will be discussed and modelled in Chapter 9.

7.2.4 There are finitely many basic *psychological operations* on propositions.

Comments. The psychological propositional operations are broadly of two kinds: *viz.*, *cognitive* and *emotive*. Emotive operations I will not deal with, beyond noting that for the cognitive operations to work — and for psychological processes to ensue in action — there must be some emotions providing the 'motive force'. Concerning the basic *cognitive* operations, Locke's position is that they are *epistemic* and *evaluative*; that is, they are operations by which the mind confirms or disconfirms (or, in general, determines the truth-values of) propositions; and all such epistemic-evaluative operations are founded, as the opening quotation in Section 7.1 indicates, on the mind's faculty of *discerning* the semantic identity or non-identity between ideas.

Although discerning underlies all evaluative operations (including such as the mind may deploy to confirm propositions without demonstrative certainty), there is one sort of evaluation especially relevant to our concerns: namely, evaluation by which the mind acquires the Lockian *intuitive* and *demonstrative knowledge*. To take the most trivial illustration, the mind can determine, with certainty, that any two tokens of the same type of idea are semantically identical, and any two tokens of different types of idea are semantically non-identical; hence it can determine that, for example, $a=a$ is true, $a=b$ is false, *etc.* (We should recall that a and b stand for — or, in our formal model of the mental code, *are* — semantically simple empirical ideas such as **blue** or **yellow**; see Section 7.2.1. So in saying that the mind can determine that $a=a$ is true, $a=b$ is false, *etc.*, I say that it can determine, by semantic analysis alone, the truth-values of such propositions as **blue is identical to blue**, **blue is identical to yellow**, and so forth; and it can make such determinations because it discerns the semantic identity, or meaning, of each simple idea, thus knowing with certainty that any two tokens of the idea **blue** have the same meaning, and any two tokens of the ideas **blue** and **yellow**, respectively, have different meanings. More on ideational meaning in Section 7.3, and on epistemic evaluation solely from meaning in Section 7.5.)

To spell out the Lockian epistemic-evaluative operations for intuitive and demonstrative knowledge — and their significance for our notions of modal properties of propositions, including conditional propositions and arguments — is one of the goals of this chapter. However, as in the case of the generative mechanisms for complex ideas and propositions, I will depart from Locke's *empiricism* — which holds that these operations do not require any non-empirical or other ideas in order to function — in assuming that the operations are *idea-laden*, comprising several classes of *a priori* ideas; in particular, ideas of *modal properties* and relations of propositions, ideas of *epistemic values* (*i.e.*, truth-values), and ideas which may be regarded as *individual constants* (with some reservations, to be mentioned later).

I will explain what these ideas are, and how they work in the epistemic-evaluative operations, in Section 7.5. For the moment, we should note only that the *a priori* faculty consists of at least three components: the two generative mechanisms for complex ideas and for propositions; and the epistemic-evaluative operations on propositions, by which the mind acquires intuitive and demonstrative knowledge. The last component completes the *a priori* faculty as a faculty of *judgement*, enabling the mind to assess the logical modality of any proposition (*pace* the moderns who believe in undecidability), and to determine the epistemic values of such propositions as may be known with intuitive or demonstrative certainty.

7.2.5 There is a generative mechanism for the production of *complex psychological operations* on propositions.

Comments. Arguably, the complex cognitive operations include expecting, doubting, recalling, concluding, and so forth; on the emotive side, there are, arguably, fearing, hoping, aspiring, *etc.* We need not worry about the exact identity of any of these operations, except in as much as we want a general notion of a *propositional mental state*, and a *propositional mental process*:

(i) a particular *mental state*, such as believing that bats are birds, is an instance of the belief-operation on a proposition (*i.e.*, a token of the mental sentence) that means that bats are birds;

(ii) a particular *mental process* (*e.g.*, thought, deliberation) is a causal sequence of instances of such operations.

So much for the five key assumptions underlying CTM. To summarise our model of CTM *thus far*, we have a representational mental code which is akin to the usual language of monadic $QT_=$. One of the differences is that our basis of simple symbols is *finite*, and a basic symbol may be tokened either as a grammatical *subject*, or as an *object*, or as a *predicate*. But the most notable differences will concern the semantics for the code, as we shall see in the next section. Further, our model differs from Locke's version of CTM, in that our basis of empirical (or *a posteriori*) ideas does not exhaust the class of semantically simple ideas: there are other, relational simple ideas (*viz.*, the idea of identity); and the generative operations for complex ideas and for propositions, as well as the epistemic-evaluative operations on propositions, also involve *a priori*, non-empirical simple and complex ideas. The *a priori* ideas laden in the epistemic-evaluative operations will be introduced in Section 7.5; and other such ideas will be defined in Sections 7.6–7.7, and in Chapters 9 and 10.

7.3 The Classical Theory of Representation

The following clauses, SP_1–SP_3, express the semantical principles implicit in CTM:

SP_1 Semantics is *term-based*: *i.e.*, the primary bearers of meaning are mental terms, or ideas; and the meaning of a complex mental symbol, idea or proposition, is built from the meanings of the simple constituents of the symbol.

SP_2 The meaning of a symbol, simple or complex, consists in that the symbol *denotes*, or *represents*, a certain *universal*: that is, a *property, relation*, or — in the case of propositions — a *state of affairs*; in other words, meanings are not (sets of) particular objects, but relations of denotation, or representation, between a symbol and a universal.

SP_3 The universals denoted, or *denotata*, are *nominal*: that is, it is the mind, rather than the environment, which determines the identity of the property or relation or state of affairs represented by a symbol; which is to say, one cannot get to understand the *real* or *noumenal* nature of the *denotatum* of a symbol merely by knowing the symbol's meaning.

There are several exceptions to principle SP_3. For Descartes, simple ideas of divine perfections (more generally, simple ideas the mind needs to contemplate divine perfections) represent *real*, mind-independent properties. For Locke, simple ideas of *primary qualities* — for instance, ideas of extension, duration, solidity, existence, unity, *etc.*, in the case of physical objects, and of consciousness, representation, duration, existence, unity, *etc.*, in the case of mental objects — represent, and are resemblances of, *real* properties. For Kant, all ideas, empirical or non-empirical, represent *nominal* rather than *real* or *noumenal* properties. Locke's view is, more precisely, that all simple ideas represent essences which are *both real and nominal*; but simple ideas of primary qualities are, in addition, *veridical* and *resemblances*, whereas simple ideas of secondary qualities are not resemblances, though still veridical, in that they signify some *real* existence from which the original tokenings of the ideas were derived, and which is causally sufficient to occasion the ideas. (See, *e.g.*, (II, VIII, 14–15; III, IV, 2–3).)

In summary, the common formula of the semantics of CTM is that the meaning of a symbol — whether simple or complex, empirical or *a priori* — consists in the relation of denotation (representation) between the symbol and a certain nominal universal. I will express this by writing: $\text{М}(\#) =_{\text{df}} \text{Д}(\#, [\#])$, where # is any symbol, [#] is the universal represented by the symbol, and М and Д are the Cyrillic counterparts of the letters "M" and "D", standing for "meaning" and "denoting", respectively. We shall now apply this precept to our model of the mental code, as it has been set out so far in Sections 7.2.1–7.2.3.

7.3.1 Basic ideas.

Each basic empirical idea represents a *mode*; that is, a monadic, non-relational, empirical property. In other words:

$\mathfrak{Z}$ $\text{М}(\alpha) =_{\text{df}} \text{Д}(\alpha, [\alpha])$, where α is any basic idea (which may be tokened as a grammatical subject, object or predicate), and $[\alpha]$ is the empirical mode represented.

7.3.2 Complex ideas, and the *a priori* ideas generating them.

There are four groups of symbols falling under this head: the simple *a priori* ideas $\neg$ and $\wedge$; the complex *a priori* ideas $\vee$, $\supset$, $\equiv$, *etc.*; the simple *a priori* predicate $=$; and the complex empirical ideas generated by the operations of the *a priori* ideas (or *a priori* operations).

7.3.2.1 *The simple* a priori *ideas* $\neg$ *and* $\wedge$.

$\mathfrak{R}_{\neg}$ $\text{М}(\neg) =_{\text{df}} \text{Д}(\neg, [\neg])$, where $[\neg]$ is an epistemic property of a proposition of the form $\neg\alpha$, such that $\neg\alpha$ is true *iff* α is not true, *iff* it is not the case that α (with α any proposition).

$\mathfrak{R}_{\wedge}$ $\text{М}(\wedge) =_{\text{df}} \text{Д}(\wedge, [\wedge])$, where $[\wedge]$ is an epistemic property of a proposition of the form $(\alpha \wedge \beta)$, such that $(\alpha \wedge \beta)$ is true *iff* both α is true and β is true, *iff* it is the case that α and it is the case that β (with α, β any propositions).

Comments. The (non-trivial) difference between CTM-based semantics and the standard *truth-conditional* semantics for these terms is that the usual definition of a truth-function does not, in the former, by itself specify the meaning of any truth-function idea, only the identity of the truth-function denoted; the meaning consists in that the idea denotes the function.

Also, the way the *functions* are set out does not matter; any convenient method, such as truth-tabular, would do equally well. I have chosen to describe $[\neg]$ and $[\wedge]$ as *epistemic* properties of propositions of such and such a form, consisting in that the propositions are true *iff* ..., *etc.*, to emphasise the epistemic character of the functions (in that they are functions of epistemic values), and to express the clauses in a way suitable for the epistemic-evaluative operations on propositions, to be described in Section 7.5. In general, we shall see in this section and in Section 7.5 that, in this model of CTM, whilst basic ideas denote *empirical* properties, all *a priori* ideas, with the exception of $=$, denote *epistemic* properties.

7.3.2.2 *The complex* a priori *ideas* $\vee$, $\supset$, *etc.*

$\mathfrak{R}_{\vee}$ $\mathbb{M}(\vee) =_{\mathrm{df}} \mathbb{D}(\vee, [\neg(\neg\alpha \wedge \neg\beta)])$;

$\mathfrak{R}_{\supset}$ $\mathbb{M}(\supset) =_{\mathrm{df}} \mathbb{D}(\supset, [\neg(\alpha \wedge \neg\beta)])$, with α, β any propositions.

Comments. Note that although the functions $[\neg(\neg\alpha \wedge \neg\beta)]$ and $[\neg(\alpha \wedge \neg\beta)]$ are specified *contextually* — as they must be, since they are relational properties — the semantics for the ideas representing the functions are *term-based*, in accord with Principle SP_1.

7.3.2.3 *The simple* a priori *predicate* $=$.

$\mathfrak{R}_{=}$ $\mathbb{M}(=) =_{\mathrm{df}} \mathbb{D}(=, [=])$, where $[=]$ is a relation between any two items ζ and η, whether particular or universal, such that ζ is identical to η.

Comments. Since $[=]$ cannot be defined in terms of properties represented by other ideas available in the model code, we take $=$ as semantically simple. This implies that the mind has only to *think* the idea, without any analysis, to understand its semantic content. However, to be thought, the idea must be tokened together with some subject and object ideas, in a mental state or process, and therefore cannot be understood in isolation. Thus although the *meaning* of $=$ does not depend on the meanings of other ideas, *understanding* it does depend on understanding other ideas.

7.3.2.4 *Complex empirical ideas.*

Complex empirical ideas are constructs from basic ideas, generated by operations of *a priori* ideas; their meanings are therefore defined by the meanings of the basic and the *a priori* ideas. I will set out clauses for complex empirical ideas which are negations and conjunctions of basic ideas; the meanings of more complex empirical ideas will be definable analogously.

$\mathbb{M}(\neg\Upsilon\phi) =_{\mathrm{df}} \mathbb{D}(\neg\Upsilon\phi, [\neg\Upsilon\phi])$, where $[\neg\Upsilon\phi]$ is an empirical property compounded of $[\Upsilon\phi]$ and $[\neg]$, such that it is not the case that ϕ is Υ (with Υ a basic idea and ϕ a variable).

$\mathbb{M}(\Upsilon\phi \wedge \Psi\omega) =_{\mathrm{df}} \mathbb{\Pi}(\Upsilon\phi \wedge \Psi\omega, [\Upsilon\phi \wedge \Psi\omega])$, where $[\Upsilon\phi \wedge \Psi\omega]$ is an empirical property compounded of $[\Upsilon\phi]$, $[\Psi\omega]$, and $[\wedge]$, such that it is the case ϕ is Υ and it is the case that ω is Ψ (with Υ and Ψ any basic ideas, and ϕ and ω any variables).

Comments. Informally, we may read the clauses as saying that the meaning of $\neg\Upsilon\phi$ consists in its denoting the property of not being Υ, and the meaning of $(\Upsilon\phi \wedge \Psi\omega)$ consists in its denoting the property of ϕ's being Υ and ω's being Ψ.

7.3.3 Propositions, and the *a priori* ideas generating them.

We have three kinds of symbol to deal with in this section: the simple *a priori* idea of indefinite quantification, Σ; the complex *a priori* idea of universality, A; and propositions generated by the operations of the *a priori* ideas.

7.3.3.1 *The simple* a priori *idea* Σ.

$\mathfrak{R}_{\Sigma}$ $\mathbb{M}(\Sigma) =_{\mathrm{df}} \mathbb{\Pi}(\Sigma, [\Sigma])$, where $[\Sigma]$ is an epistemic property of a proposition of the form $(\Sigma\delta)\Gamma\delta$, such that $(\Sigma\delta)\Gamma\delta$ is true *iff* some item, in the mind's domain of epistemic evaluation of the proposition, partakes of $[\Gamma]$ (with δ a variable, and $\Gamma\delta$ any predicate-token, whether simple or complex, in which all occurrences of δ are free, and all occurrences of other variables, if there are any, bound).

Comments. The restriction to the mind's domain of epistemic evaluation is included in the clause since the determination of epistemic value is, in the case of contingent propositions, not independent from the environment in which the propositions are tokened. For example, when a mind considers the proposition **some bats are in the orchard**, whether the proposition is true or false depends on the environment wherein the mind is embedded in an epistemically relevant way.

I will take it that every proposition has a *non-empty* domain of epistemic evaluation. The very fact that a proposition is tokened *in a mental system* ensures that something does exist in the mind's domain of evaluation for the proposition: *viz.*, the mind itself, with its symbolic states and processes. (This is the Cartesian point that when I think a proposition, I can be certain of the existence of at least one thing, namely, myself; to make it very clear, I can be certain of the existence of myself, as a mind, since the proposition, as a token of a mental sentence, must be tokened in a symbolic mental system, that is, myself; in general, there could not be a proposition without a mind which thinks it, and hence every proposition must have a non-empty domain of epistemic evaluation. This makes the proposition **nothing exists** necessarily false; clearly, though, it does not make every negative existential proposition necessarily false; nor does it make any *meaningless*.)

7.3.3.2 *The complex* a priori *idea* A.

$\mathfrak{R}_A$ $\mathbb{M}(\mathrm{A}) =_{\mathrm{df}} \text{Д}(\mathrm{A}, [\neg(\Sigma\delta)\neg\Gamma\delta])$, with δ and $\Gamma\delta$ as in $\mathfrak{R}_\Sigma$.

7.3.3.3 *Propositions.*

Like complex empirical ideas, propositions are generated from the empirical basis by operations of the *a priori* ideas; accordingly, their meanings are defined by the meanings of their empirical and *a priori* constituents. We have three classes to consider: propositions of identity between any two basic ideas; quantified propositions; and compounds by truth-functional ideas.

7.3.3.3.1 *Identity propositions.*

$\mathfrak{R}_{IP}$ $\mathbb{M}(\alpha=\beta) =_{\mathrm{df}} \text{Д}(\alpha=\beta, [\alpha=\beta])$, where $[\alpha=\beta]$ is a *state of affairs*, composed of $[\alpha]$, $[\beta]$, and $[=]$, such that $[\alpha]$ is identical to $[\beta]$ (with α and β any basic ideas).

Comments. In Section 7.5.1, we shall introduce and define the meaning of individual constants, as *a priori* ideas belonging to the epistemic-evaluative operations on propositions; and we shall allow also of identity propositions involving the constants.

7.3.3.3.2 *Quantified propositions.*

$\mathfrak{R}_{QP}$ $\mathbb{M}((\Sigma\delta)\Gamma\delta) =_{\mathrm{df}} \text{Д}((\Sigma\delta)\Gamma\delta, [(\Sigma\delta)\Gamma\delta])$, where $[(\Sigma\delta)\Gamma\delta]$ is a *state of affairs*, composed of $[\Sigma]$ and $[\Gamma]$, such that some item, in the mind's domain of epistemic evaluation of $(\Sigma\delta)\Gamma\delta$, partakes of $[\Gamma]$ (with δ a variable, and $\Gamma\delta$ any predicate-token, simple or complex, in which all occurrences of δ are free and all occurrences of other variables, if there are any, bound).

Comments. It is worthwhile to emphasise again that states of affairs, *qua* universals represented by propositional mental symbols, are *nominal* rather than *real* or *noumenal* universals; in other words, the identity of a state of affairs is determined by the proposition which represents it — more generally, by the mind and its symbolic system — rather than by the environment or *noumenal* world. This is because each *simple* idea denotes a nominal universal; and *complex* symbols, including propositions, denote universals which are constructs from universals denoted by the simple constituents of the symbols. As regards the state of affairs $[(\Sigma\delta)\Gamma\delta]$, the mere fact that the proposition $(\Sigma\delta)\Gamma\delta$ is a complex symbol guarantees that $[(\Sigma\delta)\Gamma\delta]$ is nominal, since its identity is fixed by the mind's own construction from the simple constituents of $(\Sigma\delta)\Gamma\delta$. This in turn guarantees that $(\Sigma\delta)\Gamma\delta$ retains its semantic identity (*i.e.*, meaning) independently of how the world is, or whether or not $[(\Sigma\delta)\Gamma\delta]$ is, or ever has been, instantiated in the environment.

7.3.3.3.3 *Compound propositions.*

$\mathfrak{R}_{NP}$ $\mathbb{M}(\neg\alpha) =_{\mathrm{df}} \text{Д}(\neg\alpha, [\neg\alpha])$, where $[\neg\alpha]$ is a *state of affairs*, composed of $[\neg]$ and $[\alpha]$, such that it is not the case that α (with α any proposition).

$\mathbb{M}(\alpha \wedge \beta) =_{\text{df}} \mathbb{\Pi}(\alpha \wedge \beta, [\alpha \wedge \beta])$, where $[\alpha \wedge \beta]$ is a *state of affairs*, composed of $[\alpha]$, $[\beta]$, and $[\wedge]$, such that it is the case that α and it is the case that β (with α, β any propositions).

$\mathbb{M}(\alpha \vee \beta) =_{\text{df}} \mathbb{M}(\neg(\neg\alpha \wedge \neg\beta))$.

$\mathbb{M}(\alpha \supset \beta) =_{\text{df}} \mathbb{M}(\neg(\alpha \wedge \neg\beta))$; and so on for other compounds.

This exhausts the class of ideas and propositions belonging to the model code, except for the *a priori* ideas to be defined in Sections 7.5–7.7, and in Chapters 9–10. In the next section, we shall look at Locke's position concerning the modes of epistemic evaluation of propositions, which the mind uses to acquire knowledge on the grounds of its symbolic code; in particular, those modes of evaluation whereby it acquires knowledge solely by ideational analysis and independently of any empirical matters of fact.

7.4 Aspects of the Classical Theory of Knowledge

I will start with a review of Locke's epistemology, sketching briefly his account of intuitive and demonstrative knowledge, of 'sensitive knowledge', and what he calls "right judgement". Then I will focus in detail on the account of intuitive and demonstrative knowledge, and show that Locke envisioned, but never brought to fruition, a new kind of 'Logick and Critick' proceeding from the account. Lastly, I will extract several metalogical principles underlying the account of intuitive and demonstrative knowledge, and formulate definitions of basic logical properties and relations of propositions, according to the principles.

7.4.1 Locke on epistemic evaluation.

Locke distinguishes roughly four modes of epistemic evaluation, or determination of truth-value, of propositions.

Firstly, he takes it that: "'Tis the first Act of the Mind, when it has any ... *Ideas* at all, to perceive its *Ideas*, and ... to know each what it is, and thereby also to perceive their difference, and that one is not another" (IV, I, 4). This allows the mind to evaluate certain propositions solely by its *immediate perception* of the semantic 'agreement or disagreement' among the constituent ideas comprising the propositions. For example, **blue is not yellow** is evaluable by an immediate perception of the meaning of the constituent ideas; or again, **three is more than two, and equal to one and two** is evaluable "at the first sight of the *Ideas* together, by bare *Intuition...*" (IV, II, 1). Locke calls the knowledge acquired by this mode of epistemic evaluation "intuitive knowledge": for "...in this, the Mind is at no pains of proving or examining, but perceives the Truth, as the Eye doth light, only by being directed toward it" (*ibid.*).

I will spell out Locke's position on intuitive knowledge in detail in Section 7.4.2, and again in Section 7.5; at present, I wish to stress only that

Locke does not claim that intuitively known propositions are understood and evaluated as 'semantical units', without any analysis whatever. He says rather that the mind perceives the semantic identity of the constituent ideas of the propositions, and hence is able to determine the epistemic values of the propositions 'at first sight of the ideas together'. The opposite claim, that such propositions are known as units, without any analysis, entails that they are psychologically innate, precisely the position Locke stands against.

Secondly, some propositions — though still evaluable from the mind's perception of the 'agreement or disagreement' among ideas — depend nevertheless for their evaluation on *intervening proofs*, or *semantic demonstrations*. The proposition **the three angles of a triangle are equal to two right ones** is evaluable, according to Locke, not from the mind's *immediate* perception of the semantic agreement between the ideas comprising the proposition, but rather from the its intuitive knowledge, or immediate perception, of a number of semantic connections between the proposition's constituent terms and various *intermediary ideas*. Locke calls the knowledge acquired by this mode of epistemic evaluation "demonstrative knowledge", or sometimes "rational knowledge" (IV, II, 2–13).

Thirdly, Locke allows that some propositions are evaluable solely from the mind's *sensory perception* of the 'agreement or disagreement' between, on the one hand, the ideas whereof the propositions consist, and, on the other hand, "*the particular existence of finite Beings* without us" (*ibid.*, 14). For example, the proposition **this marigold is yellow**, thought of a particular yellow marigold on a particular occasion, is evaluable solely from the mind's sensory perception of the agreement between the constituent ideas comprising the proposition, and 'the particular existence of the finite beings' referred to by the ideas. Locke calls the knowledge so acquired "sensitive knowledge".

Lastly, *most* propositions Locke regards as evaluable neither from the mind's intuitive or demonstrative perception of the 'agreement or disagreement' among ideas, nor from its sensory perception of the 'agreement or disagreement' between ideas and the 'finite beings without us', but "only as they more or less agree to Truths that are established in our Minds, and as they hold proportion to other parts of our Knowledge and Observation" (IV, XVI, 12). Notably, scientific hypotheses purporting to represent the imperceptible causes of observable events are confirmable only in larger theoretical contexts; not one-by-one, in isolation from the theories they occur in. Locke calls this kind of epistemic evaluation "right judgement" (or "belief", "opinion") (IV, XIV, 4).

7.4.2 Locke's prevision of a new 'Logick and Critick'.

I will now turn to elaborate on the notions of intuitive and demonstrative knowledge. One of the best sources to this end is Chapter VII of Book IV, where Locke discusses the nature of our knowledge of logical maxims or axioms, and their role in the demonstrative sciences. The chapter opens thus:

> There are a sort of Propositions, which under the name of *Maxims* and *Axioms*, have passed for Principles of Science: and because they are *self-evident*, have been supposed innate, without that any Body (that I know) ever went about to shew the reason and foundation of their clearness or cogency. It may however be worth while, to enquire into the reason of their evidence, and see whether it be peculiar to them alone, and also examine how far they influence and govern our other Knowledge. (IV, VII, 1)

Locke raises three questions here: the first concerns the 'reason of evidence' of logical maxims or axioms; the second, whether the certainty of such propositions is 'peculiar to them alone'; the third, how far they 'influence and govern our other Knowledge'. The rest of the chapter is organised to answer these questions: in particular, Section 2 answers the first, Sections 3–7 answer the second, and Sections 8–11 the third question. I will set out the answers in the same order.

Locke says that axioms or maxims the mind knows to be true with certainty since it can perceive — intuitively, without involved reasoning — the agreement or disagreement among the semantically simple ideas whereof such propositions consist. This intuitive perception of the agreement or disagreement among ideas is possible only because each idea has a *clear and distinct* — that is, *well-defined* and *unique* — semantic identity (*cf.* (II, XXIX, *passim*)). The semantic clarity and distinctness of the simple ideas "is so absolutely necessary, that without it there could be no Knowledge, no Reasoning, no Imagination, no distinct Thoughts at all. By this the Mind clearly and infallibly perceives each *Idea* to agree with it self, and to be what it is; and all distinct *Ideas* to disagree, *i.e.* the one not to be the other" (IV, I, 4).

In short, the evidence for the logical truth of an axiom or maxim is drawn *solely from its semantically simple constituent ideas*; and this is possible because each simple idea has a clear and distinct semantic identity, so that the mind can discern infallibly that any two tokens of the same type of idea are semantically identical, and any two tokens of different types of idea are non-identical.

As regards the second question, whether this kind of evidence is 'peculiar to axioms or maxims alone', Locke's answer is negative; he says that it is *common to all intuitively and* hence, by extension, *demonstratively true propositions* (since the latter are provable *via* intermediate ideas, where each connection is known with intuitive certainty). That blue is not yellow, that two bodies cannot be in the same place at the same time, or even that a hill is higher than its valley, are equally certain, and the nature of their evidence is the same as that of, say, the principle of non-contradiction.

Concerning the third question, 'how far logical axioms or maxims influence and govern our other knowledge', Locke's view is that *each logically true proposition*, including such as are sometimes taken as axioms, *is entirely independent of any other proposition*, whether logically true, false, or contingent; in other words, it does not provide evidence to, or receive evidence from, any other proposition whatever. This is potentially the most controversial feature of Locke's view. We shall do well to look at it closely.

Here is how Locke recapitulates the issue:

> ... let us consider, what *influence* these received *Maxims* have, upon the other parts of our Knowledge. The Rules established in the Schools, that all Reasonings are *ex præcognitis, et præconcessis*, seem to lay the foundation of all other Knowledge, in these Maxims, and to suppose them to be *præcognita*; whereby, I think, is meant these two things: First, That these Axioms, are those Truths that are first known to the Mind; and, secondly, That upon them, the other parts of our Knowledge depend. (IV, VII, 8)

Locke sets out to deny these two scholastic claims, discussing — among other aspects of the claims — axiomatic reasoning in arithmetic, whereby distinct propositions are derived one from another, with some of the most general ones serving as axioms. Locke's objection to such reasoning is that a *particular* arithmetical proposition, such as "$1+2 = 3$", the mind knows with certainty before any general axioms, and therefore must have some *other* way of proving it, this way "being nothing else but the perception it has of the agreement, or disagreement of its *Ideas*..." (IV, VII, 9). Locke then proceeds to strengthen his objection, referring to all propositions knowable with certainty:

> ...this is true of them, that they are all known by their native Evidence, are wholly independent, receive no Light, nor are capable of any proof one from another; much less the more particular, from the more general; or the more simple, from the more compounded: the more simple, and less abstract, being the most familiar, and the easier and earlier apprehended... [T]he Evidence and *Certainty* of all such Propositions is in this, That a Man sees the same *Idea* to be the same *Idea*, and infallibly perceives two different *Ideas* to be different *Ideas*... For a Man cannot confound the *Ideas* in his Mind, which he has distinct: That would be to have them confused and distinct at the same time, which is a contradiction ... (IV, VII, 10)

For the Schoolmen of Locke's day, this position was difficult to accept (see Section 11, *ibid.*). One of its startling consequences is that the demonstrative sciences — logic, mathematics, and perhaps other (Locke regarded geometry and also the foundations of ethics and theology as demonstrative) — could be done equally well or better by pure ideational analysis, and without any axiomatic reasoning, or any *proposition-based* reasoning. A transition to such demonstrative sciences could be contemplated only if the project of CTM, of spelling out in detail the structure and function of the mind — as a system of symbols and symbolic operations — were brought to fruition. Locke does not claim to have accomplished this, or even to have formulated the project fully in its logical facets. But he expresses his confidence that such a project is feasible, and that it would "afford us another sort of Logick and Critick, than what we have been hitherto acquainted with" (IV, XXI, 4). He refers to this 'Logick and Critick' at times as 'the original way of Knowledge' (IV, XVII, 4), or — borrowing a phrase from Richard Hooker — as the 'right helps of Art'; and he leaves the task of working out the details to others, "to cast about for new Discoveries, and to seek in their own Thoughts, for those *right Helps of Art...*" (IV, XVII, 7).

7.4.3 Metalogical principles and definitions.

The account of intuitive and demonstrative knowledge can be put concisely by the following three clauses, MLP_1-MLP_3, which I will take as metalogical principles for my model of CTM. Three points of nomenclature before setting out the clauses. Firstly, a proposition will be said to have such and such a modal or logical property — *e.g.*, to be logically true — just in case it is *epistemically evaluable*, by the mind's cognitive-evaluative operations, *as true solely from the meanings of its constituent simple* a priori *ideas* (and regardless of any other evidence). The term "evaluable" will be read as saying that there *is*, not that there *may be*, a way of evaluating the proposition solely from the meanings of its constituent simple ideas. Secondly, the expression "solely from the meanings of its constituent simple ideas" I will abbreviate by the scholastic phrase "*ex terminis*", which will indicate that the evaluation is to proceed *from the ends* of semantic analysis; that is, from the meanings of the semantically simple, ultimate constituent ideas, not just from the lexical constituents of the proposition. (The term "*ex terminis*" is used by Leibniz (1981: IV, VII, 1), in commenting on Locke's position concerning our knowledge of logical maxims or axioms. Leibniz points out that "scholastic philosophers have said that such propositions are evident *ex terminis* — from the terms — as soon as they are understood". However, Leibniz does not regard such evidence as capable of constituting a proof. He thinks that a proof, even in the mind's representational code, must be proposition-based rather than term-based: *i.e.*, such that some propositions — the axioms — are indemonstrable, and the mind knows them with certainty *innately* and *immediately*, without any analysis; whereas all other propositions knowable with certainty are derivable from the axioms. This contrasts with Locke's view that the mind has no innate propositional knowledge, and that no propositions knowable with certainty are epistemically prior to any other propositions.) Thirdly, the expression "solely from the meanings of its constituent simple *a priori* ideas" I will abbreviate by the phrase "*ex terminis* and *a priori*", or sometimes, where the context disambiguates it, merely by the phrase "*a priori*". (We shall see in Chapter 9 that *ex terminis* evaluation may be *a posteriori* and therefore contingent, relying on *a posteriori* or empirical ideas, in addition to *a priori* ideas; again, we shall see in Section 7.7 that *a priori* evaluation may be deductive and proposition-based, rather than term-based, and therefore not strictly *ex terminis*; however, any deductive, proposition-based evaluation will be reducible to *ex terminis* evaluation.)

Some further comments regarding my usage of the term "*ex terminis* and *a priori*": I just said that "*ex terminis* and *a priori*" is short for "[evaluable] solely from the meanings of [a proposition's] *constituent* simple *a priori* ideas", and this is how the term is to be understood in this book; for here I will deal only with uncomplicated examples of propositions, the *ex terminis* and *a priori* evaluation of which need not rely on any other simple *a priori* ideas except those which are *constituents* of the propositions. However, in a general treatment of the *ex terminis* and *a priori* method of proof, I will want to say that such an evaluation of a proposition may draw upon, in addition to the proposition's constituent simple *a priori* ideas, the mind's total resources of semantically simple *a priori* ideas; in other words, that the term "*ex terminis* and *a priori*" is short for "[evaluable] solely from the meanings of the mind's simple *a priori* ideas" (dropping the restriction to the *constituent* simple *a priori* ideas of the proposition under evaluation). The key aspect of the *ex terminis* and *a priori* evaluation of a proposition is therefore not that it is done from the proposition's constituent simple *a priori* ideas, but rather that it is done from simple *a priori* ideas as such; and these may be any simple *a priori* ideas belonging to the symbolic system of the mind. Still, to repeat, all examples of *ex terminis* and *a priori* evaluation to be discussed in this book are of the restricted sort, relying solely on the constituent simple *a priori* ideas of the proposition being evaluated.

With these provisions, the metalogical principles of CTM may be expressed thus:

MLP$_1$ Each simple idea, whether empirical (*a posteriori*) or *a priori*, has a clear and distinct semantic identity (*i.e.*, meaning).

MLP$_2$ The logical truth of a proposition consists in that the proposition is evaluable as true *ex terminis* and *a priori*, by the cognitive-evaluative operations of the mind.

MLP$_3$ Each logically true proposition is evidentially independent of any other logically true proposition.

I will finish this section with a list of informal definitions of elementary modal properties and relations of propositions, given according to the principles of CTM (and without reference to possible worlds or any such entities commonly invoked in contemporary accounts).

A proposition is logically (analytically, necessarily):

(1) true *iff* evaluable as true *ex terminis* and *a priori*;
(2) false *iff* evaluable as false *ex terminis* and *a priori*;
(3) contingent *iff* evaluable neither as true nor as false *ex terminis* and *a priori*;

(4) possibly true *iff* not evaluable as false *ex terminis* and *a priori*;

(5) possibly false *iff* not evaluable as true *ex terminis* and *a priori*.

Two or more propositions are logically:

(6) compatible *iff* their conjunction is possibly true;

(7) incompatible *iff* their conjunction is logically false;

and similarly, *mutatis mutandis*, for other such relations. Of special interest is the relation of logical *indifference*; in the light of MLP_3, two propositions are logically:

(8) indifferent *iff* at least one is either logically true or logically false, or the conjunction of each one of them with the negation of the other is contingent.

We shall now return to our model of CTM.

7.5 Modal Properties in the Model Code

This section will resume the topic of Section 7.2.4, which is to define the cognitive-evaluative operations by which the mind confirms or disconfirms such propositions as may be known by *ex terminis*, *a priori* analysis alone, with intuitive or demonstrative certainty. Section 7.5.1 will set out the analytic evaluative operations in general, for the model of CTM. Section 7.5.2 will make clear, by working out a couple of examples of *ex terminis* and *a priori* analysis, just how the operations function. Section 7.5.3 will discuss some consequences of the *a priori* analytic method for problems concerning psychological innateness and *a priori* knowledge. (As mentioned earlier, *a priori* synthetic evaluation, and *ex terminis a posteriori* synthesis, will be dealt with in Chapter 9.)

7.5.1 Analytic evaluation in the model of CTM.

I pointed out in Sections 7.2.4 and 7.4.2 that, according to Locke, all epistemic evaluation — logical evaluation in particular — is founded on the mind's capacity to *discern* the semantic identity of each simple idea, and hence ascertain the identity of any two tokens of the same type of idea, and the non-identity of any two tokens of different types of idea. I will accept Locke's insight as a starting point.

However, as in the case of his *generative* operations for complex ideas and for propositions, Locke takes it that the *cognitive-evaluative* operations, which rest on the mind's faculty of discerning, are not *idea-laden*: that is, although they are operations on propositions constructed from empirical ideas, they require no non-empirical ideas in order to function. Accordingly, Locke does not distinguish between empirical and non-empirical ideas, either as concerns the generative operations for complex ideas and for propositions, or as concerns the evaluative operations on propositions.

Contrary to Locke, I will assume that the mind's evaluative operations are *idea-laden*, and regard the ideas involved in the operations, like those of the generative mechanisms, as *a priori*; in other words, as ideas which are non-empirical, and which become psychologically active, in the epistemic evaluation of propositions, only when some propositions are formed in the mind in response to experiential stimuli. *A priori* ideas are thus active in the formation and processing of complex mental symbols in at least three stages: firstly, there are *a priori* ideas that generate *complex ideas* from the basic empirical ideas; secondly, there are *a priori* ideas that generate *propositions* from the complex as well as simple ideas; thirdly, there are *a priori* ideas active in the *evaluative operations* on propositions.

I will distinguish five evaluative operations: namely, assuming an epistemic value of the proposition under evaluation, α; inferring the values of the atomic constituent propositions of α; discerning the semantic identity of the atomic propositions in terms of their semantically simple constituent ideas; judging the epistemic value of α; and judging the logical modality of α. Clauses AO_1–AO_5 will explain the operations in detail, including the *a priori* ideas laden in them. I will ask the reader to treat these explanations patiently. It may not be immediately clear how the operations are to be integrated in an evaluative *procedure*; but whatever obscurity may remain after setting out the clauses will be removed by way of examples in Section 7.5.2. (The symbol "AO" is short for "analytic operation"; in Chapter 9, we shall match the analytic operations with corresponding synthetic operations and a synthetic evaluative procedure.)

AO_1 *Assume* the proposition under evaluation has a certain *epistemic value*, either *true* or *false*.

Comments. Using the symbols τ for the value *true*, φ for the value *false*, and '**AssumeV**(...) = ...' for the operation of assuming the value, we can write AO_1 as '**AssumeV**(α) = τ (or φ)', where α is any proposition. The symbol '**AssumeV**(...) = ...' stands for a propositional operation rather than an idea of any sort; however, the symbols τ and φ stand for — or *are*, in our model of CTM — *a priori* ideas, and as such need to be given formal semantical clauses, with φ defined in terms of τ:

$\mathfrak{R}_\tau$ $\mathbb{M}(\tau) =_{\text{df}} \mathbb{I}(\tau, [\tau])$, where $[\tau]$ is the epistemic value *true* of a proposition α, such that α is true *iff* it is the case that α.

$\mathfrak{R}_\varphi$ $\mathbb{M}(\varphi) =_{\text{df}} \mathbb{I}(\varphi, [\varphi])$, where $[\varphi]$ is the epistemic value *false* of a proposition α, such that α is false *iff* $\neg\alpha$ is true, *iff* it is not the case that α.

AO_2 *Infer the epistemic value* of each *atomic constituent* of the proposition under evaluation, by the semantical clauses $\mathfrak{R}_\#$.

Comments. I will write AO_2 as '**InferV**(α) = τ/φ ($*$, $\mathfrak{R}_\#$)', where α is any constituent proposition of the proposition under evaluation, $*$ stands for the number(s) of the line(s) from which the inference is drawn, and $\mathfrak{R}_\#$ indicates the semantical clause(s) by which it is drawn. The purpose of AO_2 is to resolve the proposition under evaluation into its *atomic constituents* and to assign a truth-value to each, given the assumption made by AO_1 (and making subordinate assumptions if needed). Atomic constituents are propositions of the form $K\lambda$, or $\zeta=\eta$, or the negations thereof, where K is any *basic* predicate idea, λ any *individual constant*, and ζ and η are either any basic ideas, or any individual constants. An individual constant I will regard as an *a priori* idea laden in, or belonging to, the operation of inferring, tokened when the inference is drawn by $\mathfrak{R}_{\Sigma}$ from a proposition of the form $(\Sigma\delta)\Gamma\delta$, and denoting an item, whether particular or universal, which the mind *supposes*, on the strength of the assumption made by AO_1, to exist in the domain of epistemic evaluation of $(\Sigma\delta)\Gamma\delta$, and to partake of $[\Gamma]$. In other words, an individual constant, λ, is an *a priori* idea, tokened in the operation $\mathbf{InferV}(\Gamma\lambda) = \tau\ (* , \mathfrak{R}_{\Sigma})$, where $*$ assigns the truth-value $[\tau]$ to a proposition of the form $(\Sigma\delta)\Gamma\delta$; and the meaning of λ consists in its representing an item $[\lambda]$ (supposed, on the strength of AO_1, to exist in the domain of evaluation of $(\Sigma\delta)\Gamma\delta$, and to partake of $[\Gamma]$). I will take it that the mind can form an unlimited number of individual constants, and use the lower-case letters *u, v, w* — with numerical subscripts if necessary — to represent them in the model code; strictly speaking, I should take it that the mind can form a finite number of *semantically simple* individual constants, and form further such ideas by some generative mechanism; but I need not complicate the issue at this stage. (What I call "individual constant" is not quite the same as an individual constant of Russellian logic; in the latter, individual constants are *proper names* standing for individual objects, and these are distinguished from predicates standing for properties defining sets of objects; in contrast, my individual constants are not proper names, but rather something like *pseudonyms* for individual objects, since the ideas *u, v, w, etc.*, are used only *within a process of evaluation*, and may be re-used within another process of evaluation; thus, instead of being proper names, these ideas are used in a certain symbolic procedure, or — so to speak — *for a certain literary sake*, as pseudonyms are characteristically used.)

The semantical clauses for these ideas, and for the atomic propositions in which they occur, are as follows:

$\mathfrak{R}_{IC}$ $\mathbb{M}(\lambda) =_{df} \text{Д}(\lambda, [\lambda])$, where λ is any individual constant, and $[\lambda]$ is the item, particular or universal, represented.

$\mathfrak{R}_{AP1}$ $\mathbb{M}(K\lambda) =_{df} \text{Д}(K\lambda, [K\lambda])$, where $[K\lambda]$ is a *state of affairs*, composed of $[K]$ and $[\lambda]$, such that $[\lambda]$ partakes of the property $[K]$ (with K any *basic* empirical predicate, and λ any individual constant).

$\mathfrak{R}_{AP2}$ $\mathbb{M}(\zeta=\eta) =_{df} \text{Д}(\zeta=\eta, [\zeta=\eta])$, where $[\zeta=\eta]$ is a *state of affairs*, composed of $[\zeta]$, $[\eta]$, and $[=]$, such that $[\zeta]$ is identical to $[\eta]$ (with ζ, η any individual constants).

AO$_3$ *Discern the meaning* of each atomic proposition in terms of its simple constituent ideas, and determine whether each idea has a clear and distinct semantic identity, given the assignment of epistemic value to the atomic proposition.

Comments. I will use the phrase '**DiscernM**(#) $=_{\mathbf{df}}$ *Д*(#, [#])' for the operation of discerning, where # is either a semantically simple constituent idea of an atomic proposition, or the atomic proposition itself; and I will form a list of such phrases, to include every simple idea occurring in each atomic proposition. The purpose of discerning is to find:

(a) exactly what the atomic propositions say, *in terms of their semantically simple constituent ideas*;

(b) whether each simple idea of an atomic proposition has a clear and distinct — that is, well-defined and unique — semantic identity, *given the assignment of epistemic value* to the atomic proposition by AO$_2$.

Whenever there is a semantic conflict between the simple constituent ideas of one or more atomic propositions — given the assignments of epistemic values to the propositions — I will indicate it by writing '**DiscernM**($\#_1$, $\#_2$, ...)**NC&D** ($*_1$, $*_2$,...)', where $\#_1$, $\#_2$,... are the semantically simple ideas involved in the conflict, '**NC&D**' stands for "not clear and distinct", and $*_1$, $*_2$,... refer to the lines in which the ideas occur. If no semantic conflict is discerned, I will write '**DiscernM**($\#_1$, $\#_2$,...)**C&D** ($*_1$, $*_2$,...)', where $\#_1$, $\#_2$,... list all the semantically simple constituent ideas of the proposition under evaluation, the symbol '**C&D**' stands for "clear and distinct", and $*_1$, $*_2$,... refer to the lines in which the simple ideas occur in atomic propositions.

AO$_4$ *Judge the epistemic value* of the proposition under evaluation.

Comments. I will write AO$_4$ as '**JudgeV**(α) $= \tau/\varphi$', where α is the proposition under evaluation.

The operation is to work thus:

If, in the scope of the assumption made by AO$_1$ (and each of its subordinate assumptions, where there are any), the *meanings* of some simple ideas are discerned to be not clear and distinct, then the assignment of *epistemic value* made by AO$_1$ is judged to be the opposite: in other words, supposing **AssumeV**(α) $= \varphi$, AO$_4$ will yield **JudgeV**(α) $= \tau$; supposing **AssumeV**(α) $= \tau$, it will yield **JudgeV**(α) $= \varphi$.

If, on the contrary, in the scope of **AssumeV**(α) $= \tau$ (or at least one of its subordinate assumptions, where there are any), the *meanings* of all simple ideas are discerned to be clear and distinct, and if the same holds also for **AssumeV**(α) $= \varphi$, then the judgement of epistemic value is *suspended*; which I will write as '**JudgeV**(α) $= \nu$', where ν stands not for a truth-value, but rather indicates that AO$_4$ is unable to determine the truth-value.

Briefly, AO_4 may be set out as follows, where $\#_1$, $\#_2$,... are the semantically simple constituent ideas of the atomic propositions obtained by the operation AO_2:

(i) if **AssumeV**(α) $= \varphi$ and **DiscernM**($\#_1$, $\#_2$, ...)**NC&D** in all subordinate assumptions, then **JudgeV**(α) $= \tau$;

(ii) if **AssumeV**(α) $= \tau$ and **DiscernM**($\#_1$, $\#_2$, ...)**NC&D** in all subordinate assumptions, then **JudgeV**(α) $= \varphi$;

(iii) if **AssumeV**(α) $= \tau$ and **DiscernM**($\#_1$, $\#_2$,...)**C&D** in at least one subordinate assumption, and if the same holds also for **AssumeV**(α) $= \varphi$, then **JudgeV**(α) $= \nu$.

AO_5 *Judge the logical modality* of the proposition under evaluation.

Comments. I will write AO_5 as '**JudgeMOD**(...α) $= \tau/\varphi$', where α is the proposition under evaluation, and '...' makes room for the *a priori* ideas of logical modalities laden in the operation: *viz.*, $\Box$ for necessity, $\Diamond$ for possibility, and ∇ for contingency.

The operation may be set out thus:
if **JudgeV**(α) $= \tau$, then **JudgeMOD**($\Box\alpha$) $= \tau$;
if **JudgeV**(α) $= \varphi$, then **JudgeMOD**($\Diamond\alpha$) $= \varphi$;
if **JudgeV**(α) $= \nu$, then **JudgeMOD**($\nabla\alpha$) $= \tau$.

The semantical clauses for the *a priori* ideas of modal properties of propositions will be as follows, with $\Diamond$ and ∇ defined in terms of $\Box$ (and with no reference to possible worlds, *etc.*):

$\mathfrak{R}_{\Box}$ $\mathbb{M}(\Box) =_{df} \mathbb{I}(\Box, [\Box])$, where $[\Box]$ is an epistemic property of a proposition of the form $\Box\alpha$, such that $\Box\alpha$ is true *iff* α is evaluable as true *ex terminis* and *a priori*, by the evaluative operations AO_1–AO_4 (with α any proposition).

$\mathfrak{R}_{\Diamond}$ $\mathbb{M}(\Diamond) =_{df} \mathbb{I}(\Diamond, [\Diamond])$, where $[\Diamond]$ is an epistemic property of a proposition of the form $\Diamond\alpha$, such that $\Diamond\alpha$ is true *iff* $\neg\Box\neg\alpha$ is true, *iff* α is not evaluable as false *ex terminis* and *a priori*, by the evaluative operations AO_1–AO_4 (with α any proposition).

$\mathfrak{R}_{\nabla}$ $\mathbb{M}(\nabla) =_{df} \mathbb{I}(\nabla, [\nabla])$, where $[\nabla]$ is an epistemic property of a proposition of the form $\nabla\alpha$, such that $\nabla\alpha$ is true *iff* $(\neg\Box\neg\alpha \wedge \neg\Box\alpha)$ is true, *iff* α is evaluable neither as true nor as false *ex terminis* and *a priori*, by the evaluative operations AO_1–AO_4 (with α any proposition).

I will assume that once the mind has acquired these ideas, in the evaluative operation AO_5, it is able to use them in the generative operations for complex ideas and for propositions, to generate complex ideas such as $\Box(Fx \vee \neg Fx)$, $\Diamond Gy$, and so forth, propositions such as $(\mathrm{A}x)\Box(Fx \vee \neg Fx)$, $\Box(\Sigma y)\Diamond Gy$, *etc.*

Let us now summarise AO_1-AO_5, before turning to examples. **AssumeV** assigns an arbitrary epistemic value to the proposition under evaluation. **InferV** resolves the proposition into atomic propositions, and assigns to each an epistemic value in accord with **AssumeV**. **DiscernM** sorts out just what the atomic propositions say, in terms of the semantically simple ideas comprising them; and whether each idea has a clear and distinct semantic identity, given the assignments of epistemic value to the propositions. **JudgeV** determines the epistemic value of the proposition under evaluation, depending on whether or not the simple ideas retain clear and distinct semantic identity, given the assignments of value to the atomic propositions. Lastly, **JudgeMOD** assesses the modal status of the proposition, so that whenever **JudgeV** determines the value of the proposition as so-and-so, the proposition is so-and-so with logical necessity; and whenever **JudgeV** cannot determine the value, the proposition is logically contingent.

The gist of the operations is to examine whether the assumed epistemic value of the proposition under evaluation accords with the semantic identity of the simple ideas comprising the proposition; if it does not, then the assignment of value must give way to the identity of meaning. The simple ideas, to put it in a Lockian way, 'the mind can neither make nor change'; they are the rock foundation of the mind, and constrain not only which propositions the mind can form and think, but also which can be bearers only of truth, which only of falsehood, and which of either truth or falsehood. The logical modality of a proposition thus consists in the constraint the semantic identity of the constituent simple ideas of the proposition imposes on the proposition's epistemic value. This is the notion of modality Locke expresses in his definition of logical falsehood: "*Contrary to Reason* are such Propositions, as are inconsistent with, or irreconcilable to our clear and distinct *Ideas*" (IV, XVII, 23). Again, Descartes relies on this notion of modality when he defines the term "possible truth" to "mean what everyone commonly means, namely 'whatever does not conflict with our human concepts'..." (1984: 107).

The apparent circularity in Locke's definition I have resolved by my operations AO_1–AO_5. It is worth stressing that AO_1–AO_5 are indeed *operations*, not symbols or ideas belonging to the code; as such, they are not bearers of any semantic properties, and can and should be specified, in the final account, syntactically. Hence although the modal *ideas* $\Box$, $\Diamond$, and ∇ are defined in terms of **InferV**, and although the operations **JudgeV** and **JudgeMOD** I have described using clauses of the form "if..., then...", the definitions of $\Box$, $\Diamond$, and ∇ are not circular, since they involve no unspecified modal *ideas*.

Lastly, I wish to reiterate that AO_1–AO_5 are *analytic* operations, allowing the mind (in this model of CTM) to acquire *a priori* analytic knowledge: that is, to know with certainty solely by ideational analysis. We shall see in Chapter 9 that this is not the only way the mind can know with certainty; rational ideational synthesis will also deliver certainty. In general, any necessarily true proposition will be provable either by rational *ex terminis* and *a priori* analysis, or by rational *ex terminis* and *a priori* synthesis (contrary to Kant's notion of the analytic-synthetic distinction but, as we shall see, in agreement with Descartes and others).

I will now turn to work out a couple of examples of *ex terminis* and *a priori* analysis, taking it as an occasion to answer potential queries, and to remove any remaining obscurity in the account.

7.5.2 The trifling propositions $a=a$ and $(\mathrm{A}x)(Fx \supset Fx)$.

Consider first the trifling proposition $a=a$.

1. **AssumeV**($a=a$) $= \varphi$.
2. **InferV**($\neg a=a$) $= \tau$ (1, $\mathfrak{R}_{\varphi}$).
3. **DiscernM**(a) $=_{\mathrm{df}}$ ⌷(a, [a]), where a is a basic idea, and [a] is the empirical mode represented (2, $\mathfrak{Z}$).
4. **DiscernM**($=$) $=_{\mathrm{df}}$ ⌷($=$, [$=$]), where [$=$] is a relation between any items ζ and η, such that ζ is identical to η (2, $\mathfrak{R}_{=}$).
5. **DiscernM**($\neg$) $=_{\mathrm{df}}$ ⌷($\neg$, [$\neg$]), where [$\neg$] is an epistemic property of a proposition of the form $\neg\alpha$, such that $\neg\alpha$ is true *iff* α is not true, *iff* it is not the case that α (with α any proposition) (2, $\mathfrak{R}_{\neg}$).
6. **DiscernM**(τ) $=_{\mathrm{df}}$ ⌷(τ, [τ]), where [τ] is the epistemic value *true* of a proposition α, such that α is true *iff* it is the case that α (2, $\mathfrak{R}_{\tau}$).
7. **DiscernM**($\neg a=a$) $=_{\mathrm{df}}$ ⌷($\neg a=a$, [$\neg a=a$]), where [$\neg a=a$] is a state of affairs, composed of [a], [$=$] and [$\neg$], such that it is not the case that [a] is identical to [a] (2–5, $\mathfrak{R}_{\mathrm{IP}}$, $\mathfrak{R}_{\mathrm{NP}}$).
8. **DiscernM**(a, $=$, $\neg$, τ)**NC&D** (2–7).
9. **JudgeV**($a=a$) $= \tau$ (1–8).
10. **JudgeMOD**($\Box a=a$) $= \tau$ (1–9).

Comments. I do not attempt to automate the procedure by an algorithm, but I think it is obvious that this could be done; especially Line 8, the discerning of not clear and distinct meaning of the simple ideas, would need spelling out, to bring it forth that the ideas a, $=$, $\neg$, and τ cannot simultaneously have, under the assumption of epistemic value of Line 1, clear and distinct semantic identity.

Turn now to another trifling proposition, $(\mathrm{A}x)(Fx \supset Fx)$.

1. **AssumeV**($(\mathrm{A}x)(Fx \supset Fx)$) $= \varphi$.
2. **InferV**($(\Sigma x)\neg(Fx \supset Fx)$) $= \tau$ (1, $\mathfrak{R}_{\varphi}$, $\mathfrak{R}_{\mathrm{A}}$, $\mathfrak{R}_{\neg}$).
3. **InferV**($\neg(Fu \supset Fu)$ $= \tau$ (2, $\mathfrak{R}_{\Sigma}$).
4. **InferV**(Fu) $= \tau$ (3, $\mathfrak{R}_{\neg}$, $\mathfrak{R}_{\supset}$).
5. **InferV**($\neg Fu$) $= \tau$ (3, $\mathfrak{R}_{\neg}$, $\mathfrak{R}_{\supset}$).
6. **DiscernM**(F) $=_{\mathrm{df}}$ ⌷(F, [F]), where F is a basic predicate, and [F] is the empirical mode represented (4, 5, $\mathfrak{Z}$).
7. **DiscernM**(u) $=_{\mathrm{df}}$ ⌷(u, [u]), where u is an individual constant, and [u] is the item represented (4, 5, $\mathfrak{R}_{\mathrm{IC}}$).

8. **DiscernM**$(\neg) =_{\mathrm{df}} \text{Д}(\neg, [\neg])$, where $[\neg]$ is an epistemic property of a proposition of the form $\neg\alpha$, such that $\neg\alpha$ is true *iff* α is not true, *iff* it is not the case that α (with α any proposition) $(5, \Re_{\neg})$.
9. **DiscernM**$(\tau) =_{\mathrm{df}} \text{Д}(\tau, [\tau])$, where $[\tau]$ is the epistemic value *true* of a proposition α, such that α is true *iff* it is the case that α $(4, 5, \Re_{\tau})$.
10. **DiscernM**$(Fu) =_{\mathrm{df}} \text{Д}(Fu, [Fu])$, where $[Fu]$ is a state of affairs, composed of $[F]$ and $[u]$, such that it is the case that $[u]$ partakes of $[F]$ $(4, 6, 7, 9, \Re_{\mathrm{AP}_1})$.
11. **DiscernM**$(\neg Fu) =_{\mathrm{df}} \text{Д}(\neg Fu, [\neg Fu])$, where $[\neg Fu]$ is a state of affairs, composed of $[F]$, $[u]$ and $[\neg]$, such that it is not the case that $[u]$ partakes of $[F]$ $(5\text{–}9, \Re_{\mathrm{AP}_1}, \Re_{\mathrm{NP}})$.
12. **DiscernM**$(F, u, \neg, \tau)$**NC&D** (4–11).
13. **JudgeV**$((\mathrm{A}x)(Fx \supset Fx)) = \tau$ (1–12).
14. **JudgeMOD**$(\Box(\mathrm{A}x)(Fx \supset Fx)) = \tau$ (1–13).

Comments. It is important to bear in mind that each simple idea, whether empirical or *a priori*, is clear and distinct when it is considered by itself, in isolation; further, no simple ideas can become not clear and distinct merely by being combined into a proposition, whatever the proposition might be. The only way to render some ideas not clear and distinct is, firstly, to combine them into a certain proposition — such as $(\mathrm{A}x)(Fx \supset Fx)$ or $a=a$ — and, secondly, to assign a certain epistemic value to the proposition (in our examples, the value φ). In short, it is not ideas as such, nor propositions, which can be rendered not clear and distinct, but *ideas* combined into a *proposition* under a certain assumption of *epistemic value*.

Notice also that **DiscernM** works only on the constituents of the atomic propositions and on the simple idea τ. Hence the only ideas that can enter into a semantic conflict in **DiscernM** are the finitely many basic empirical ideas, the simple ideas τ, $\neg$, and $=$, and individual constants. This indicates that the operation of discerning could be abbreviated, and the evaluative procedure much simplified.

However, the procedure cannot be so abbreviated as to stop simply at the level of inferring the epistemic values of the atomic propositions (Lines 4 and 5); that would be to rely solely on the Aristotelian principle of non-contradiction, which is proposition-based rather than term-based, and which rests on the *uniqueness of truth-value* rather than *uniqueness of meaning*. The point of the logic of CTM is precisely that it resolves the Aristotelian criterion of the uniqueness of truth-value of a proposition to the criterion of the uniqueness of meaning, under an assumption of truth-value, of the semantically simple ideas comprising the proposition (a criterion which may be regarded as Platonic, in agreement with the many Platonic features implicit in CTM, and with the general anti-Aristotelian push common to all early-modern CTM theorists, including Locke).

7.5.3 Comments on innate and *a priori* knowledge.

Locke calls propositions such as $a=a$, $(\mathrm{A}x)x=x$, $(\mathrm{A}x)(Fx \supset Fx)$ "trifling propositions", since they "add no Light to our Understandings, bring no increase to our Knowledge" (IV, VIII, 1). The proposition $(\mathrm{A}x)x=x$ Locke often mentions as a case of *putative innate knowledge*; that is, as a proposition which allegedly could not allow of any demonstration or logical analysis, and hence — on the psychological side — had to be taken as known innately and immediately, as a semantical and logical unit. This is the doctrine Locke sets out to refute; and his chief argument is that even though such propositions are known *intuitively*, in that the mind is in 'no pains of proving or examining' in evaluating them as true, they are nevertheless not known as units, without any analysis whatever; rather, the mind knows the clear and distinct semantic identity of the simple constituent ideas of the propositions, and therefore is able to determine that the ideas semantically 'agree one with another'. Such 'agreement or disagreement' I have sought to capture in my model of CTM; and it has this significance for the classical philosophy of mind, that it explains how the mind could demonstrate any analytically true proposition, no matter how trifling, so that the proposition need not be supposed innate but rather known by a proof, however trivial the proof may seem.

As to innate *ideas* rather than propositions, Locke's first concern is to show that no ideas need be supposed psychologically active prior to and independently of some or other experiences. Locke thereby rejects the Cartesian view that the soul can think regardless of any experiences (*e.g.*, in a disembodied state, as Descartes might have it). Another of his concerns is that each simple idea should answer *adequately* and *veridically* to some aspect of the environment (whether external or internal). These are the foremost reasons why Locke denies the distinction between empirical and non-empirical ideas, taking all simple ideas to be empirical; for he fears that admitting *non-empirical* ideas would be tantamount to admitting *non-veridical* representations in the very constitution of the mind.

I will suggest, in closing this section, that Locke could have afforded himself of the useful and correct distinction between empirical and non-empirical or *a priori* ideas, such as I have drawn in my model of CTM, even holding that all mentation depends on and follows after some experiences (as Kant held), and even holding (contrary to Kant) that the non-empirical ideas have an adequate and veridical application to the world as it is; which would have allowed him to avoid thorough empiricism and make use of the theoretical benefits of rationalism. These benefits are already apparent in this model of CTM, notwithstanding its simplicity; but as we gradually enrich the model, adding ideas of space, time, of causal relations, of number, of the self — and also, should we follow the classical theorists, of the good and bad, *etc.* — the distinction between empirical and *a priori* ideas will become still more prominent and indispensable.

7.6 The Complex Ideas of Implication

There is a sense in which only in this section we shall come to the proper business of this chapter, which is the logic implicit in the Classical Theory of Mind; for our exposition of the logic would not be complete — in fact, would hardly have begun — without a treatment of the complex *a priori* and *a posteriori* ideas of implication, and of valid deduction. We shall treat of ideas of implication in this section, and of that of valid deduction in the next; and we shall require all the material covered in Sections 7.2–7.5 to do it.

I will call propositions of the form "if α were true, then β *would* be true" "strong conditionals", and those of the form "if α were true, then β *might* be true" "weak conditionals"; and I will distinguish two kinds of either strong or weak conditional: *necessary conditionals*, which are epistemically evaluable solely by *ex terminis* and *a priori* analysis; and *contingent conditionals*, the evaluation of which depends in part on *a posteriori* ideas. The propositional relation represented by "if ..., then ...", whether in a strong or a weak conditional, I will call "implication"; and, in this model of CTM, I will use the symbol $\rightarrow$ for the idea of strong necessary implication, $\mapsto$ for the idea of weak necessary implication, $\Rightarrow$ for the idea of strong contingent implication, and $\rightharpoonup$ for the idea of weak contingent implication.

The purpose of this section is to define the semantic identity of these ideas, in accord with the metalogical principles of CTM. But firstly, we shall need to consider the significance of these principles for the cognitive operations of *assuming the antecedent* and *implying the consequent* in a *true conditional*. This will be the topic of Section 7.6.1; Section 7.6.2 will formulate the four semantical clauses, and Section 7.6.3 will conclude with some comments concerning the normativity of complex ideas, ideas of implication specifically, in CTM-based logic.

7.6.1 Assuming and implying in a true conditional.

We have seen in Sections 7.4.2–7.4.3 that, according to CTM, all necessary propositions, true or false, are logically independent, and hence neither imply nor are implied by any other proposition; they are "known by their native Evidence, are wholly independent, receive no Light, nor are capable of any proof one from another; much less the more particular, from the more general; or the more simple, from the more compounded" (*op. cit.*, Section 7.4.2). This is a *very* extraordinary aspect of CTM-based logic, so it will be worthwhile to look at it in detail, with regard to the issue of assuming the antecedent and implying the consequent of a true conditional.

7.6.1.1 Assuming a logical falsehood.

A proposition is logically false just in case the mind has a way of evaluating it as false *ex terminis* and *a priori*. Therefore the mind cannot, *without violating the principle of clear and distinct ideas*, assume a logical falsehood as true; for in assuming it as true, the mind has the capacity to work out, by its cognitive-evaluative operations, that the constituent ideas of the proposition do not have, *under the assumption*, clear and distinct semantic identity, and hence that the proposition cannot be true. It follows that — though any proposition may be assumed as true, insofar as assuming is merely a cognitive-evaluative operation, ultimately to be defined syntactically — not any proposition may be assumed as the antecedent of a *true conditional*. In particular, no conditional with a logically false antecedent can be true; for in assuming such an antecedent as true, the mind is in a position to discern that the simple constituent ideas of the antecedent do not have, under the assumption, clear and distinct semantic identity, and so judge that the conditional as a whole cannot be true.

Let us now work over the same claim, and the same argument, using a simple example. Take $\neg a=a$ as the antecedent, and β as any consequent, of the conditional $\neg a=a \rightarrow \beta$ (equivalently, use $\rightarrowtail$, $\Rightarrow$, or $\rightharpoonup$ instead of $\rightarrow$). Even *before* knowing the precise semantics for the idea $\rightarrow$ (or $\rightarrowtail$, $\Rightarrow$, $\rightharpoonup$), we can tell, by the principles of CTM, that the conditional cannot be true. For in assuming the antecedent $\neg a=a$ as true, we are able to show — by an *ex terminis*, *a priori* analysis analogous to that whereby we proved in Section 7.5.2 that $a=a$ is logically true — that the simple constituent ideas of the antecedent (*viz.*, $\neg$, a, and $=$) do not have, under the assumption, clear and distinct semantic identity, so that the antecedent $\neg a=a$, and hence the conditional $\neg a=a \rightarrow \beta$ as a whole, must be false. In general, we can show likewise that since the mind cannot assume a logical falsehood as true *without violating the principle of clear and distinct ideas*, no conditional with a logically false antecedent can be true.

7.6.1.2 Implying a logical falsehood.

Similarly, no conditional with a logically false consequent can be true. For such a consequent is evaluable as false *ex terminis* and *a priori*, and therefore the mind cannot — without violating the principle of clear and distinct ideas — imply it as true, whatever antecedent is assumed. Thus $\beta \rightarrow \neg a=a$, where β is any antecedent, must be false, since the consequent $\neg a=a$ cannot be implied as true whilst preserving the clear and distinct semantic identity of the simple ideas $\neg$, a, and $=$.

7.6.1.3 Assuming a logical truth.

In the mental code (as opposed to public-language discourse, where generally some aspects of the thought a speaker intends to convey are left unexpressed or ambiguous), the truth of the antecedent of a true *strong* conditional, whether necessary or contingent, is *sufficient* for the truth of the consequent; and the truth of the consequent is *necessary* for the truth of the antecedent. When the consequent is false, so that the necessary condition for the truth of the antecedent is not satisfied, the antecedent is also false. In other words, the conditional $\alpha \rightarrow \beta$ (or $\alpha \Rightarrow \beta$) is *truth-functionally equivalent* to $\neg\beta \rightarrow \neg\alpha$ (or $\neg\beta \Rightarrow \neg\alpha$); contraposition holds, in the mental code, for strong conditionals. Further, given that contraposition holds for the strong conditionals, it holds *a fortiori* for weak conditionals; that is, $\alpha \rightarrowtail \beta$ and $\alpha \rightharpoonup \beta$ are truth-functionally equivalent to, respectively, $\neg\beta \rightarrowtail \neg\alpha$ and $\neg\beta \rightharpoonup \neg\alpha$.

It follows that, by the principles of CTM, no conditional proposition with a logically true antecedent can be true. For such a conditional — for instance, $a{=}a \rightarrow \beta$, where β is any consequent — is truth-functionally equivalent to $\neg\beta \rightarrow \neg a{=}a$, which we know to be false, since its consequent cannot be implied as true without violating the clear and distinct semantic identity of the simple ideas $\neg$, a, and $=$. (*Cf.* Section 7.6.1.2.)

7.6.1.4 Implying a logical truth.

Analogously, it follows that no conditional with a logically true consequent can be true. For such a conditional — *e.g.*, $\beta \rightarrow a{=}a$, where β is any antecedent — is truth-functionally equivalent to $\neg a{=}a \rightarrow \neg\beta$, which we know to be false, since the antecedent $\neg a{=}a$ cannot be assumed as true without violating the clear and distinct semantic identity of the simple ideas $\neg$, a, and $=$. (*Cf.* 7.6.1.1.)

Putting 7.6.1.1–7.6.1.4 together, we may judge that a necessary condition for the truth of a conditional proposition is that both its antecedent and its consequent be logically contingent.

This necessary condition is clearly enunciated by Locke, as I have shown in Section 7.4.2. Versions of it have appeared in modern Analytic Philosophy with Strawson, Von Wright, Geach, Smiley, and Watling. But none of the modern versions have contained cogent defences of the position, and all are baffled by the apparent conflict between, on the one hand, the view that a necessary proposition neither implies nor is implied by any other necessary proposition, and, on the other hand, the common practice of asserting conditionals with necessary antecedents and consequents, and — as in *reductio ad absurdum* arguments, or in axiomatic proofs — of implying necessary conclusions on the grounds of necessary premisses. I will now turn to show that this conflict is indeed only apparent, and does not at all contravene the logic of CTM.

The view that a necessary proposition neither implies nor is implied by any other proposition holds, as I have argued in Sections 7.6.1.1–7.6.1.4, for the *mental code*; in the mental code, any necessary proposition is in principle provable, *independently of any other proposition*, by the epistemic-evaluative operations of the mind. I have illustrated how this is done in Section 7.5.2, using as examples the trifling propositions $a{=}a$ and $(\mathrm{A}x)(Fx \supset Fx)$.

However, when an *ex terminis*, *a priori* proof of a proposition in the mental code is, so to speak, *translated* into and communicated in a *public code* — that is, in a public-language discourse (whether formalised, as in mathematics and formal logic, or 'ordinary') — we are inclined to express the sequence of the epistemic-evaluative operations on the proposition as if it were a sequence of if-then conditionals, such that the consequent of each conditional is *logically dependent* on the antecedent. For example, in the *ex terminis*, *a priori* proof of $(\mathrm{A}x)(Fx \supset Fx)$, the mind may begin with the operation **AssumeV**$((\mathrm{A}x)(Fx \supset Fx)) = \varphi$, and continue with the operation **InferV**$((\Sigma x)\neg(Fx \supset Fx)) = \tau$. But when it communicates this step in a public language, it is wont to express it by saying that *if* $\neg(\mathrm{A}x)(Fx \supset Fx)$, *then* $(\Sigma x)\neg(Fx \supset Fx)$. This, though it seems innocuous, is misleading, since it gives the impression that the mind *implies* the necessary consequent from the necessary antecedent, which in fact, in its *ex terminis* and *a priori* method of proof of $(\mathrm{A}x)(Fx \supset Fx)$, it does not. Strictly speaking, then, the mind should endeavour to mirror as closely as it can, in its public expressions of the *a priori* proof, the procedure of the proof, which involves only the epistemic-evaluative operations on the proposition and on its constituent ideas, with no implications from necessary antecedents to necessary consequents. This is what I have done in my own public-language formulation of the model of CTM, and of the *ex terminis* and *a priori* method of proof in the model.

In general, the apparent conflict between the classical mentalistic view that a necessary proposition neither implies nor is implied by any other proposition, and the common public practice of asserting conditionals with necessary antecedents and consequents, and of drawing necessary conclusions from necessary premisses, disappears when the public expressions of the mind's *ex terminis* and *a priori* method of evaluation and proof mirror — as they should, insofar as the primary function of public language is to express thought — the structure of that method.

7.6.2 The complex ideas $\rightarrow$, $\rightarrowtail$, $\Rightarrow$, and $\rightharpoonup$.

I will now turn to define the semantic identity of the ideas of implication. These are complex ideas (like $\Diamond$ and ∇) which I class as belonging to the operation of judging logical modality, AO_5. The *a priori* ideas of necessary implication are simpler than the *a posteriori* ideas of contingent implication, and will be set out first.

7.6.2.1 The complex *a priori* ideas of necessary implication.

$\mathfrak{R}_{\rightarrow}$ $\mathbb{M}(\rightarrow) =_{\mathrm{df}}$ ℿ$(\rightarrow, [\rightarrow])$, where $[\rightarrow]$ is an epistemic property of a proposition of the form $(\alpha \rightarrow \beta)$, such that $(\alpha \rightarrow \beta)$ is true *iff* $(\Box\neg(\alpha \land \neg\beta) \land \nabla\alpha \land \nabla\beta)$ is true; in other words, *iff*:

- *(a)* $(\alpha \land \neg\beta)$ is evaluable as false *ex terminis* and *a priori*,
- *(b)* α is evaluable neither as false, nor as true, *ex terminis* and *a priori*,
- *(c)* β is evaluable neither as false, nor as true, *ex terminis* and *a priori*,

by the evaluative operations AO_1–AO_4 (with α, β any propositions).

Comments. The modality $[\rightarrow]$ may be said to consist in that $(\alpha \rightarrow \beta)$ is true just in case $(\alpha \land \neg\beta)$ is contradictory, in the *ex terminis* sense, but its contradictoriness is not due to either of the conjuncts taken individually; rather, it is due to their conjunction; that is the gist of the account. This kind of account has a rather interesting recent history. Using the term "entailment" where I have used "strong necessary implication", Smiley (1959) distinguishes between accounts of entailment in terms of *logical consequence*, and those in terms of *deducibility*, or *deductive systems*. The former accounts are similar to my own. For example, Strawson (1948: 186) suggests that "[t]he expression "'*p*' entails '*q*'" is to be used to mean "'$p \supset q$' is necessary, and neither '*p*' nor '*q*' is either necessary or self-contradictory"...". Von Wright (1957: 178) agrees with Strawson's suggestion, but thinks that it "stands in need of further clarification", and that "[t]o account of its meaning *is* to account of the meaning of entailment". Von Wright then offers a clarification as follows: "*p* entails *q*, if and only if, by means of logic, it is possible to come to know the truth of $p \rightarrow q$ without coming to know the falsehood of *p* or the truth of *q*" (p. 181; with → standing for ⊃). In turn, Geach (1958) accepts Von Wright's clarification, but replaces the expression 'it is possible to...' with '*there is* a way of...', and the expression 'by means of logic' with 'there is an *a priori* way'. His account then reads thus: "*p* entails *q* if and only if there is an *a priori* way of getting to know C*pq* which is not a way of getting to know either whether *p*, or whether *q*" (p. 164; C*pq* stands for $p \supset q$). A similar view is defended by Watling (1958), who has much to say concerning the requirement of contingency.

Smiley considers his own version of this kind of account, which is designed to allow entailments such as $A \;\&\; {\sim}A \vdash A$ (where ⊢ stands for 'entails', & for ∧, and ~ for ¬), whilst guaranteeing that $A \;\&\; {\sim}A \vdash B$ holds for *not any B* whatever. The account is this: "$A_1, \ldots, A_n \vdash B$ if and only if the implication $A_1 \;\&\; \ldots \;\&\; A_n \supset B$ is a substitution instance of a tautology $A'_1 \;\&\; \ldots \;\&\; A'_n \supset B'$, such that neither $\vdash B'$ nor $\vdash {\sim}(A'_1 \;\&\; \ldots \;\&\; A'_n)$" (p. 240). Smiley thinks that, save for his account, any theory involving the requirement that the antecedent and the consequent of a true entailment be contingent has to fail, since "under it *reductio ad absurdum* argument is strictly impossible, for no premiss *can* on this theory entail a contradiction or even the two arms of a contradiction separately" (p. 244). He then proceeds to develop the latter approach, in terms of *deducibility* or *deductive systems*, which was pioneered by Ackermann (1956). This approach has become more influential in subsequent years; the former approach was framed in terms of possible worlds, but the condition that the antecedent and the consequent of a true entailment be contingent tended to wither. Anderson and Belnap (1975) deal with the problem of entailment entirely in terms of deducibility: "Since we wish to interpret "$A \rightarrow B$" as "*A* entails *B*", or "*B* is deducible from *A*," we clearly want to be able to *assert* $A \rightarrow B$ whenever there exists a deduction of *B* from *A*..." (p. 7). However, they still regard the former approach as 'plausible and interesting' (§§ 5.1.1, 15.1, 20.1).

My CTM-based account falls under "Logical Consequence" rather than "Deducibility"; and, as we shall see in Section 7.7.1, the idea of valid deduction will be accordingly defined in terms of that of strong necessary implication, not conversely. The logical-consequence approach may be further categorised as involving accounts in terms of possible worlds, and the early-modern (Lockian and Cartesian) accounts in terms of clear and distinct ideas; my proposal obviously falls into the latter category. It is notable that what Smiley took to be the greatest hurdle for the consequence approach — namely, *reductio ad absurdum* arguments — makes no trouble for that approach when it is of the latter, early-modern, non-possible-worlds-theoretic variety; I have shown how such arguments work in Section 7.5.

$\mathfrak{R}_{\rightarrowtail}$ $\mathbb{M}(\rightarrowtail) =_{\mathrm{df}} \Pi(\rightarrowtail, [\rightarrowtail])$, where $[\rightarrowtail]$ is an epistemic property of a proposition of the form $(\alpha \rightarrowtail \beta)$, such that $(\alpha \rightarrowtail \beta)$ is true *iff* $(\Diamond \neg(\alpha \wedge \neg\beta) \wedge \nabla\alpha \wedge \nabla\beta)$ is true; in other words, *iff*:

(a) $(\alpha \wedge \neg\beta)$ is not evaluable as true *ex terminis* and *a priori*,

(b) α is evaluable neither as true, nor as false, *ex terminis* and *a priori*,

(c) β is evaluable neither as true, nor as false, *ex terminis* and *a priori*,

by the evaluative operations AO_1–AO_4 (with α, β any propositions).

Comments. The modality $[\rightarrowtail]$ is that $(\alpha \rightarrowtail \beta)$ is true *iff* $(\Diamond \neg(\alpha \wedge \neg\beta) \wedge \nabla\alpha \wedge \nabla\beta)$ is true, which in turn is true just in case $((\alpha \rightarrow \beta) \vee (\nabla\neg(\alpha \wedge \neg\beta) \wedge \nabla\alpha \wedge \nabla\beta))$ is. From this we can see that $((\alpha \rightarrow \beta) \supset (\alpha \rightarrowtail \beta))$ is logically true; that is, $\Box((\alpha \rightarrow \beta) \supset (\alpha \rightarrowtail \beta))$ is true. However, $((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrowtail \beta))$ is not true, since the antecedent is either logically true or logically false, and so is the consequent. Again, $\Box((\alpha \rightarrow \beta) \supset (\neg\beta \rightarrow \neg\alpha))$ is true, but $((\alpha \rightarrow \beta) \rightarrow (\neg\beta \rightarrow \neg\alpha))$ is not.

I mentioned in Section 7.5.1 that the *a priori* ideas of logical properties — $\Box$, $\Diamond$, and ∇ — once acquired in the *evaluative* operations, may enter into the *generative* operations for complex ideas and for propositions, to form ideas such as $\Box(Fx \vee \neg Fx)$, propositions such as $\nabla(\Sigma x)Gx$, $(\mathrm{A}y)\Diamond Hy$, and so forth. Similarly, the *a priori* ideas of logical relations — *viz.*, $\rightarrow$ and $\rightarrowtail$ — once acquired, may enter into the generative operations, to form ideas such as $((Fx \wedge Gx) \rightarrow Fx)$, $(Fx \rightarrow (Fx \vee \neg Fx)$, propositions such as $(\mathrm{A}x)((Fx \wedge Gx) \rightarrow Fx)$, which will be true, $(\mathrm{A}x)(Fx \rightarrow (Fx \vee \neg Fx))$, which will be false, *etc.*

Let us now very briefly consider a form of proposition known as "the Barcan formula", $\Diamond(\Sigma x)\Psi x \supset (\Sigma x)\Diamond\Psi x$. It should be clear upon reflection that, depending on the idea Ψ, the antecedent and the consequent of a proposition of that form will be either both logically true, or both logically false. For supposing Ψ is such that $(\Sigma x)\Psi x$ is possibly true, $\Diamond(\Sigma x)\Psi x$ is necessarily true; but then also Ψ is such that Ψu, for any individual constant

u, is possibly true, and hence $(\Sigma x)\Diamond\Psi x$ is necessarily true (for recall that every proposition must have a non-empty domain of epistemic evaluation; see Section 7.3.3.1). Conversely, supposing Ψ is such that $(\Sigma x)\Psi x$ is not possibly true, $\Diamond(\Sigma x)\Psi x$ is necessarily false; but then also Ψ is such that Ψu, for any constant u, is necessarily false, and hence $(\Sigma x)\Diamond\Psi x$ is necessarily false. Therefore every proposition of the form $\Diamond(\Sigma x)\Psi x \supset (\Sigma x)\Diamond\Psi x$ is a logical truth. However, no proposition of the form $\Diamond(\Sigma x)\Psi x \rightarrow (\Sigma x)\Diamond\Psi x$ is true, since both its antecedent and its consequent are necessary, contrary to the requirement that both must be contingent. (For those who wish to be reminded, the Barcan formula raises a problem for classical quantificational modal logic, since any proposition of that form turns out logically true, under standard (S5) possible-worlds interpretation; yet, construing "if ..., then ..." as $\Box(\ldots \supset \ldots)$, the proposition says that from the mere *possibility* of there being something which is Ψ, it follows that there *is* something which is possibly Ψ; a consequence not guaranteed, intuitively, by the antecedent. CTM resolves the issue, since although all propositions of the form $\Diamond(\Sigma x)\Psi x \supset (\Sigma x)\Diamond\Psi x$ are logically true, those of the form $\Diamond(\Sigma x)\Psi x \rightarrow (\Sigma x)\Diamond\Psi x$ — *i.e.*, the claims to implication — are false.)

7.6.2.2 The complex *a posteriori* ideas of contingent implication. The account in this section will be concerned solely with *logical features of contingent implication*, aiming to show how the mind can form complex, clear and distinct ideas thereof; there will be no attempt to confront the problem of contingent implication in its *a posteriori* aspects. The purpose of the section is to point out the manner by which the mind can begin to extend its knowledge of necessary, purely *a priori* conditional propositions, to contingent *a posteriori* conditional propositions (some features of which may be knowable by *a priori* means, either by rational analysis or by rational synthesis; *cf.* Chapter 9).

The epistemic evaluation of a contingent *a posteriori* conditional will involve contingent propositions representing either natural laws, or nominal laws such as conventional regularities, or any other laws or states of affairs relevant to the implication between the antecedent and the consequent. I will refer to such propositions as "accessory propositions", and leave it open, for the present purposes, exactly what they might be. The ideas of strong and weak contingent implication ($\Rightarrow$ and $\rightarrow$, respectively) may be then defined as follows:

$\mathfrak{R}_{\Rightarrow}$ $\mathbb{M}(\Rightarrow) =_{\mathrm{df}} \mathbb{\Pi}(\Rightarrow, [\Rightarrow])$, where $[\Rightarrow]$ is an epistemic property of a proposition of the form $(\alpha \Rightarrow \beta)$, such that $(\alpha \Rightarrow \beta)$ is true *iff* $(\Box\neg(\alpha \wedge \neg\beta \wedge \Omega) \wedge \nabla(\alpha \wedge \neg\beta) \wedge \nabla\alpha \wedge \nabla\beta)$ is true; *i.e.*, *iff*:

- *(a)* $(\alpha \wedge \neg\beta \wedge \Omega)$ is evaluable as false *ex terminis* and *a priori*,
- *(b)* $(\alpha \wedge \neg\beta)$ is evaluable neither as false, nor as true, *ex terminis* and *a priori*,

(c) α is evaluable neither as false, nor as true, *ex terminis* and *a priori*,

(d) β is evaluable neither as false, nor as true, *ex terminis* and *a priori*,

by the evaluative operations AO_1–AO_4 (with α and β any propositions, and Ω a conjunction of *true* accessory propositions).

$\mathfrak{R}_{\rightharpoonup}$ $\mathbb{M}(\rightharpoonup) =_{df} \mathbb{\Pi}(\rightharpoonup, [\rightharpoonup])$, where $[\rightharpoonup]$ is an epistemic property of a proposition of the form $(\alpha \rightharpoonup \beta)$, such that $(\alpha \rightharpoonup \beta)$ is true *iff* $(\Diamond \neg(\alpha \wedge \neg\beta \wedge \Omega) \wedge \nabla(\alpha \wedge \neg\beta) \wedge \nabla\alpha \wedge \nabla\beta)$ is true; *i.e.*, *iff*:

(a) $(\alpha \wedge \neg\beta \wedge \Omega)$ is not evaluable as true *ex terminis* and *a priori*,

(b) $(\alpha \wedge \neg\beta)$ is evaluable neither as true, nor as false, *ex terminis* and *a priori*,

(c) α is evaluable neither as true, nor as false, *ex terminis* and *a priori*,

(d) β is evaluable neither as true, nor as false, *ex terminis* and *a priori*,

by the evaluative operations AO_1–AO_4 (with α and β any propositions, and Ω a conjunction of *true* accessory propositions).

Comments. The purpose of subclause *(b)* — *viz.*, $\nabla(\alpha \wedge \neg\beta)$ — is to separate the ideas of contingent implication from the ideas of necessary implication, so that the former and the latter do not overlap under any circumstances; it is to put the onus of contingent implication on Ω, as well as α and β, excluding the possibility that the onus be on α and β alone.

Some of the true accessory propositions in Ω might be other conditionals representing either natural or nominal laws, and these may or may not be evaluable with respect to still further conditionals belonging to further sets of accessory propositions. If one adopts a thoroughly *empiricist* stance — denying the distinction between empirical and non-empirical ideas, and denying the possibility of *a priori* knowledge — one should regard the ultimate law-representing conditionals as evaluable at best by empirical means, involving some mix of holistic and inductive confirmation; and one should regard the choice of ultimate law-representing conditionals as *pragmatic*. But if one adopts a *rationalist* stance — drawing a distinction between empirical and non-empirical ideas, and allowing of *a priori* knowledge — one may regard some of the ultimate law-representing conditionals as *a priori*, whether analytic or synthetic, and thus not contingent. The Classical Theory of Mind, even in Locke's version, has always adopted the latter stance, and accordingly regarded physical science, and knowledge of ultimate natural laws, as having a metaphysical foundation knowable with *a priori* certainty. (Locke would not have used these terms, but in fact he has drawn the comparable distinctions; see Section 9.2.1.)

7.6.3 Comments concerning the normativity of complex ideas. The ideas →, ↣, ⇒, and → are *complex* ideas; whether the mind forms them or not depends on its needs, social and environmental circumstances, and above all on its self-knowledge. So the mind may or may not, as a matter of fact, form these ideas. However, with respect to the metalogical principles (MLP_1–MLP_3), which I take to be implicit in the system of mind, the formation of these complex ideas is not arbitrary. I will mention two reasons.

Firstly, the component of the semantical clauses which says that a strong conditional cannot hold true unless the antecedent is incompatible with the negation of the consequent (or, in the case of weak conditionals, unless the antecedent is incompatible with the negation of the consequent, or the antecedent and the consequent are indifferent) does seem to be universal: the mind has no option but to construct its complex ideas of implication so as to include at least this much.

Secondly, insofar as the mind needs to construct these ideas with respect to the principles MLP_1–MLP_3, it cannot but include in their definitions the constraint that necessary propositions, whether true or false, neither imply nor are implied by any other propositions, and hence that the antecedent and the consequent of a true conditional must be contingent; I have shown the reasons for this constraint in Sections 7.4.2–7.4.3 and 7.6.1.1–7.6.1.4.

In general, there is an objective and universal *normative constraint* on the formation of the complex ideas of modal properties and relations, set by the clear and distinct semantic identity of the mind's simple ideas, as well as by its evaluative operations; and although the mind is able to create what complex ideas of modal properties it finds suitable or effective for its purposes, it is nevertheless bound to form such complex ideas as accord with the simple ideas and evaluative operations, if the complex ideas themselves are to be clear and distinct, free from semantic confusion and inconsistency.

7.7 The Complex *A Priori* Idea of Valid Deductive Inference

The interest of deductive reasoning lies in that the mind can proceed, by logical means alone, from premisses which, taken *in conjunction*, are contingent, and which are either assumed or known to be true, to contingent conclusions which are neither given nor yet known, thereby — to put it in Locke's idiom — *enlarging its knowledge*. This is something the mind cannot do with the ideas of implication alone; for, to know the truth of a conditional, the mind has to contemplate, by the same token, both its antecedent and its consequent, rather than deduce an unknown consequent from the antecedent.

In Section 7.7.1, I will introduce the symbol $\vdash$ to represent the complex *a priori* idea of valid deductive inference in my model of CTM, and define the semantic identity of the symbol according to the principles of CTM. In Section 7.7.2, I will consider some of Quine's objections against modal properties and relations. These objections properly belong under the heading of deductive inference, since they have to do with apparently invalid arguments which standard informal modal logic turns out as valid (as opposed to apparently false propositions, such as the Barcan formula, which it turns out as true).

7.7.1 The complex *a priori* idea $\vdash$.

The ideas of deductive inference and strong necessary implication are closely related, but not identical. Informally, an argument is deductively valid just in case there is a way of deducing the conclusion from the premisses by a finite number of applications of a certain sort of deductive rules; and the meaning of the idea of validity consists in that it denotes this epistemic property. I will assume, to define the meaning of $\vdash$ formally, that there are finitely many basic deductive rules $\rho_1, \ldots, \rho_m$, such that each rule is a *schema* of a true, strong, necessary conditional, the antecedent of which the mind regards as the premiss(es) of the rule, and the consequent as the conclusion. I need not specify just which schemata of conditional should serve as deductive rules, since my present goal is only to tell how the mind can compose, according to the principles of CTM, a clear and distinct, complex idea of deductive validity. But there are obvious examples of such schemata: $((\xi \wedge \varsigma) \rightarrow \xi)$, $(((\xi \vee \varsigma) \wedge \neg\xi) \rightarrow \varsigma)$, and so forth, where ξ and ς are *contingent* propositions, so ensuring that the antecedents and consequents constructed from them are contingent. The identity of $\vdash$ may be then defined as follows:

$\mathfrak{R}_{\vdash}$ $\mathbb{M}(\vdash) =_{df}$ Д$(\vdash, [\vdash])$, where $[\vdash]$ is the epistemic property of valid deductive inference, of an argument of the form $\alpha_1, \ldots, \alpha_n \vdash \beta$, such that $\alpha_1, \ldots, \alpha_n \vdash \beta$ is valid *iff*:

(a) there is a way of deducing β from $\alpha_1, \ldots, \alpha_n$ by a finite number of applications of the basic rules $\rho_1, \ldots, \rho_m$;

(b) $\rho_1, \ldots, \rho_m$ are such that there is a way of deducing β from $\alpha_1, \ldots, \alpha_n$ *iff* $((\alpha_1 \wedge \ldots \wedge \alpha_n) \rightarrow \beta)$ is true

(with $\alpha_1, \ldots, \alpha_n$ the finitely many premisses, and β the conclusion of the argument).

Comments. Clause *(b)* says that the system of deductive rules $\rho_1, \ldots, \rho_m$ must be sound and complete *vis-à-vis ex terminis* evaluation, in that whatever argument is provable as valid by the deductive rules must be so provable *ex terminis*, and whatever argument is provable as valid *ex terminis* must be so provable by the deductive rules. Note also that although the clause requires, implicitly, that the conjunction $(\alpha_1 \wedge \ldots \wedge \alpha_n)$ be contingent, it does not require that each of the α_j's taken individually be contingent; thus, for example, supposing $(\xi \rightarrow \varsigma)$ logically true, with ξ and ς contingent, $(\xi \rightarrow \varsigma)$, $\xi \vdash \varsigma$ is a valid argument.

In summary, valid deductive inference is like strong necessary implication; the purpose of forming the complex *a priori* idea $\vdash$, in addition to $\rightarrow$, is to enable the mind to deduce contingent consequents *not yet given or known*, something for which the idea $\rightarrow$ on its own is not suited. Were it not for this deductive aspect, there would be no point in drawing the distinction between strong necessary implication and valid inference, and in forming distinct ideas of these relations. Of the two, however, $\rightarrow$ is closer to *ex terminis* evaluation and hence more fundamental; for *ex terminis* evaluation is the final arbiter concerning the validity of any deductive argument, just as it is concerning the logical modality of any proposition.

7.7.2 Some of Quine's objections against logical modality.

Quine (1953d) raises doubts about logical modality because of arguments such as these:

(i) Necessarily 9 is greater than 7;
the number of planets = 9;
therefore necessarily the number of planets is greater than 7.

(ii) Possibly the number of planets is less than 7;
the number of planets = 9;
therefore possibly 9 is less than 7.

(iii) Necessarily if there is life on the Evening Star
then there is life on the Evening Star;
the Evening Star = the Morning Star;
therefore necessarily if there is life on the Evening
Star then there is life on the Morning Star.

According to an informal conception of standard modal logic, which Quine is working with, these intuitively invalid arguments turn out as valid, using only the substitutivity of identicals as a deductive rule. According to CTM-based logic, in contrast, no such arguments are valid, for at least two reasons.

To begin with, each of the arguments has a logically false conclusion. (For whatever modality a proposition has, it has it with logical necessity; in the case of the conclusions of *(i)*, *(ii)*, and *(iii)* — that necessarily the number of planets is greater than 7, that possibly 9 is less than 7, that necessarily if there is life on the Evening Star then there is life on the Morning Star — the modalities are clearly wrong, on intuitive grounds; so the conclusions must be regarded as logically false.) I showed earlier that the mind cannot infer such conclusions as true without violating the principle of clear and distinct ideas (see Section 7.6.1.2). This alone is sufficient to render the arguments invalid, since there can be no schemata of true strong necessary conditional that could serve as deductive rules to validate such arguments.

Secondly, each of the arguments mixes a logically true with a contingently true premiss, and draws its conclusion by substituting into the former on the basis of the latter. But although mixing necessary with contingent premisses is in itself not a sin (provided the necessary premiss is true), *substituting into a necessary premiss on the basis of a contingent identity* — that is, a contingent proposition of identity — *is a fallacy*. For the logical modality of a proposition consists in the mode of epistemic evaluation of the proposition, and therefore constrains where the evidence for the proposition may come from. In the case of the necessarily true premiss, its evidence must be drawn *ex terminis*, from the *meaning* of its constituent ideas. Hence one could not truth-preservingly substitute into the premiss, unless the replacing idea and the idea being replaced were semantically identical, or *synonymous* (*i.e.*, same in meaning); which clearly they are not, in either of the arguments.

The design of this chapter has been expository rather than polemical; so I will not engage Quine any further. But I would that my expository labour had this much effect on Quinian professors, that when they challenge the notion of logical modality, which we have inherited from the classical philosophy of mind, they confront the real thing, not — as has been usual with them — a straw proxy of their own making.

7.8 Remarks on CTM and Analytic Philosophy

At various places in the *Essay*, Locke foreshadows a 'new sort of Logick and Critick', other than either axiomatic or syllogistic logic. He tells us that this 'Logick and Critick' should rest on the mind's 'original way of Knowledge', underlying both axiomatic and syllogistic reasoning (or any conclusive reasoning whatever) and consisting in the mind's "*perception of the connexion and agreement, or disagreement and repugnancy of any of our Ideas*" (IV, I, 1). It is plausible, in fact, to read the entire *Essay* as an attempt to spell out this 'original way of Knowledge'; and although Locke does not bring the project anywhere near completion, leaving it to others to carry it farther, a diligent reading of the *Essay* will show that he comes a long way.

I will close this chapter by highlighting a few features in my elaboration on the Lockian legacy, which seem to me most characteristic of the logic implicit in it, and in CTM generally. There are three groups of criteria I will use to this end; these may be regarded broadly as concerning the *syntax, semantics*, and *epistemology* of CTM.

Among the *syntactic* features I will mention only the most comprehensive: that logical symbols are mental symbols, and that constructing a logical system requires mapping out (in practice, modelling) the system of mind, as a system of cognitive-evaluative operations on the propositions of the mental code; in turn, it requires that propositions be construed as composed of tokens of mental terms, or ideas, with complex ideas built from a finite number of basic empirical ideas by a finite number of generative operations. A major point of contention between classical theorists has been whether the basic empirical ideas exhaust the class of simple ideas, and which ideas are simple. I chose to draw a distinction between empirical and non-empirical ideas, and to regard the empirical ideas as *a posteriori*, whilst non-empirical ideas as *a priori*, supposing the latter to be psychologically active, only when some basic ideas are acquired, in the generation of complex empirical ideas and propositions, and in assessing the epistemic values and logical modalities of the propositions.

The key *semantic* features of CTM are that the fundamental bearers of meaning are mental terms rather than propositions or structures of propositions; that the meaning of a symbol consists in its denoting or representing a certain universal (not in the universal itself, let alone in a particular or set of particulars partaking of the universal); and that the universals represented are — with some qualifications — nominal: *i.e.*, fixed by the mind itself, not the environment, whether external or internal. I took each basic idea to represent an empirical mode, whilst each *a priori* idea (with the exception of the idea of identity) to represent an epistemic property, thus reflecting the special role of *a priori* ideas in epistemic evaluation.

The main *epistemic* features are, *firstly*, that the evidence for the logical truth or falsehood of a proposition is drawn solely from the proposition's semantically simple, clear and distinct constituent ideas, by certain evaluative operations, including assuming an epistemic value, inferring the values of atomic propositions, discerning the meanings of simple ideas, and judging the epistemic value and logical modality; and, *secondly*, that the criterion for the resolution of an assumption of value is not the Aristotelian principle of non-contradiction, but rather the principle of clear and distinct semantic identity of simple ideas. The foundation of logical modality, and of logic and epistemology in general, thus resides in the semantic identity of the simple ideas whereof the mind composes its complex ideas and propositions, and in the operations whereby it evaluates the propositions. For the simple ideas and the operations will not allow just any proposition to be a bearer of truth, or falsehood; some propositions they constrain to be true, some to be false, regardless of non-semantic and non-logical, empirical matters of fact. This constitutes the objective ground of logic, and puts a normative restriction on the formation of clear and distinct complex ideas, including ideas of implication and valid deductive inference.

Classical formal logic of Frege and Russell, which lies in the heart of Analytic Philosophy, may be viewed as an attempt to inject rigour into early-modern logic, and Analytic Philosophy overall as an attempt to infuse greater rigour into early-modern philosophy. Frege and later Analytic Philosophers had this much in common, that they sought an objective ground for logic outside the mind: Frege in a mind-independent realm of abstract objects; most of his followers, unhappy with this 'unscientific Platonism', and convinced that logic could not derive its objectivity from the mind, increasingly sought it in public, behavioural, overt habits and conventions (where, after much searching, they did not find it either, concluding there was no such ground). The reasons why Frege and others felt compelled to abandon mentalism, and to seek another ground for logical certainty and objectivity, had to do not so much with a clearly understood failure of mentalism, but rather with a failure to understand mentalism, as well as a need for a simplification and expediency, making the task of formalisation in logic easier to manage. In my endeavour to revive early-modern mentalism and give it a partial formal expression, I forego the beguiling benefits of purifying formal logic of its mentalistic origins, and allow it to be embroiled, once again, in the many profound problems the classical philosophy of mind encounters (the very problems the early formal logicians desired to shake off). Yet this makes logic truly a central part of philosophy, and restores to it a depth and scope which have disappeared in the modern and post-modern developments. One such problem concerns the nature of ideas, taken not only as bearers of semantic properties, and not only as components of propositions and so subjects of logical and epistemic evaluation, but also as discrete forms of consciousness. Here logic touches on the very heart of philosophy, something we ought to appreciate rather than shun.

In the next chapter, we shall take the first steps toward a theory of the nature of ideas; we shall not go nearly as far as to account for their forms of consciousness, but only for their material implementation in the brain; which, as far as first steps go, will be far enough.

Chapter 8

The Classical Theory of Mind II

8.1 The Ontology and Architecture of CTM

The Classical Theory of Mind holds that the mind is a symbolic system comprising a finite number of simple ideas, a class of generative operations for the production of complex ideas from the simples, a class of generative operations for the production of propositions from the simple and complex ideas, and a class of cognitive and emotive operations on propositions. One of the most difficult problems for CTM — one that historically underlay the conceptual objections discussed in earlier chapters — is that we do not know how to make a natural science of it: we do not know what in the brain could implement simple mental symbols, complex symbols and propositions, and propositional cognitive and emotive operations; and how such a system of mental symbols and operations could accord with what is already known of the physical constitution, organisation, and function of the brain. This problem is made further intractable by the fact, according to CTM, that each simple mental symbol is a discrete *form of consciousness*, and complex symbols are complex forms of consciousness built from the simples; for example, the simple idea **blue** has a certain syntax, a certain meaning in that it denotes the nominal essence [blue], and a certain form of consciousness, ⟦blue⟧, without which it would not be the idea it is; similarly, the proposition **blue is not yellow** has a certain syntax, a certain meaning in that it represents the nominal state of affairs [blue is not yellow], and a certain form of consciousness, ⟦blue is not yellow⟧, which is inseparable from the identity of the proposition (*qua* token of a mental sentence).

In this chapter, I will address the problem of the *material implementation* of mental symbols and the system of mental symbols posited by CTM, but without any attempt to counter the issue of what makes a simple symbol such-and-such a form of consciousness; and although I will propose a conjecture on the implementation of mental symbols in the brain, I will not thereby commit myself to materialism, taken as the doctrine that mental symbols and states, among other things, are purely physical, and that a complete account of their physical nature would be a complete account of all of their aspects. Rather, I will be concerned solely with what might be regarded as the *syntactic properties* of mental symbols; not with the syntax of this or that idea or proposition, but in general with the physical or biological kind of ideas; in other words, the problem will be, what natural kind ideas are, *insofar as they are a natural kind.*

I will discuss the problem of the nature of mental symbols in the context of the two rival research programmes in contemporary cognitive sciences: namely, *connectionism* and the *classical paradigm.* The term "connectionism" I will use in a broader than usual sense, to refer to not only the relatively recent computational models of — or inspired by — neural networks, but also, and more importantly, the cognitive theories implicit in *biological studies* of learning, memory, and simple associative processes. Such theories are guided by available neurological data, and designed to emulate the neural functions which are known to be among the mediators and determinants of behaviour, and which are considered as pertinent to cognition. These functions may be conveniently classified thus:

(i) neural *'computation'* and *'signalling'*;

(ii) the alterations in *synaptic strength* resulting from neural 'signalling';

(iii) the anatomical alterations, in *synaptic distribution*, resulting from 'signalling'.

Connectionism so understood is the leading paradigm in most, though not all, research into the biological basis of cognition, and has a history which can be traced back to the latter half of the 19th century (some of which I will survey in Section 8.5).

The term "classical paradigm" I will also use in a broader than usual sense, to include CTM, Fodor's computational mentalism, and any theories which aim to build, with or without the aid of computer modelling, a scientific psychology based on the common-sense psychology of such cognitive and emotive states as beliefs and desires. These theories construe mental states as operations — identifiable with believing, desiring, *etc.* — on tokens of the sentences of a representational mental code, or propositions, taking propositions to be constructs of tokens of mental terms, or concepts, ideas. Regarded in this sense, the classical paradigm dates back at least to Descartes, Locke and Kant.

Some connectionists hold that the two rival programmes are not incompatible, and that the classical conceptual and propositional representation will *emerge*, at a higher level of analysis, from the connectionist emulations of neural networks (for example, Smolensky 1988; Rumelhart & McClelland 1986). To the contrary, most classical theorists argue that although connectionist models may succeed in imitating the neural or some supra-neural level of analysis of the brain, they cannot succeed *as psychology* (Fodor & Pylyshyn 1988).

The most persuasive, if not always explicit, argument in favour of connectionism and against the classical paradigm has been that, concerning such fundamental mental functions as learning, memory, and information processing, the only known neural functions relevant to their biological implementation are those expressed by *(i)–(iii)*; that, so far as neuroscience tells us, there is nothing in the brain identifiable with tokens of mental terms and sentences, and operations on sentence-tokens (*e.g.*, Freeman & Skarda 1990). There is another version of this argument, often not clearly distinguished from the charge that no biological entities and processes implement the classical posits. It is assumed that the classical paradigm *requires* a certain cognitive architecture: namely, the *classical computational architecture* of the serial von Neumann machine, viewed — by analogy — as a cognitive architecture. The kernel of such an architecture is a central processor in which symbols are displayed (or tokened) and operated on, and in which computational-cognitive processes take place serially. The argument is that the brain has no central processor of that sort, and that cognitive processes in the brain are distributed and parallel rather than centralised and serial.

I will suggest that although the latter version of this argument is correct, in that the brain's cognitive architecture is almost certainly not the classical computational architecture, the former version may no longer hold good; further, the classical paradigm does not require the classical computational architecture, and therefore even the latter version looses its cogency. There are several facets of this claim, which I will make plausible in the course of this chapter.

Firstly, connectionism rests upon neurological data which are but a fraction of what neural and molecular biology have to offer to the cognitive sciences today; and the fuller range of the data allows of a new approach to cognitive theory, one that could lead to a convergence of a sort between connectionism and the classical paradigm (though not a convergence of the emergentist kind envisaged by some connectionists).

Secondly, the evidence provided by recent research in the biology of learning and memory tends to shift the burden of cognitive individuation from the level of synaptic connections to the sub-neural, molecular, *genetic level*; and this not only in the trivial sense of explaining a higher-level phenomenon by reference to entities and processes at a lower level, but in the sense of uncovering an appropriate level of analysis for cognitive individuation.

Thirdly, the evidence opens the possibility of construing mental states as genetic rather than synaptic states, which accords with the requirements and posits of the classical paradigm; in particular, it might turn out that the genetic code, which is known to function as the biological system of representation in heredity and ontogenetic development, also lies in the foundation of, and to that extent functions as, a cognitive system of representation, or *mental code*.

Lastly, though the second version of the connectionists' most persuasive argument is correct, in that the classical computational architecture is not analogous to the brain's cognitive architecture, an alternative account of cognitive architecture commends itself in the light of the genetic conjecture on the nature of mental symbols and operations, which has the desirable property of accommodating the virtues of the classical paradigm, as well as the connectionist insight that most cognitive processes are distributed in several functional areas of the brain, and many processes run in parallel.

I will structure my argument for these unusual propositions as follows: in Sections 8.2 and 8.3, I will survey some of the evidence for the role of genes in cognition; in Section 8.4, I will set out the details of the genetic conjecture, concerning the manner of implementation of mental symbols and processes at the genetic level. Finally, since similar proposals have been under discussion, on and off, throughout the past century, especially in the 1950's and 60's, I will review the most interesting historical precedents of the conjecture in Section 8.5. Readers with a background in cognitive biology need not worry about Sections 8.2 and 8.3, noting in them only the few paragraphs with links to the genetic construal of mental symbols; for these sections contain only an exposition of recent research, which should be useful for those without a previous knowledge of it. Similarly, Section 8.5 is merely historical, intended mainly for the people of fashion who will object that what I propose here is an old hat; well, I rather like old hats, as by now is perhaps obvious. The core of the chapter, requiring closest attention, is Section 8.4.

8.2 Learning, Memory, and Association in *Aplysia*

The simplest and best-known case which provides evidence for the involvement of genes in the formation, retention, and processing of cognitive states is drawn from studies in the marine snail *Aplysia*. Seen at an overt *behavioural level*, the case is as follows. *Aplysia* withdraws its respiratory organ, the gill, when its siphon (a fleshy extension through which sea-water and waste are exuded) is touched. If the siphon is touched repeatedly, *Aplysia* becomes *habituated*; that is, it ceases to withdraw the gill. A noxious stimulus, such as a shock to the head or tail, causes *sensitisation*, in that the animal again withdraws the gill when it is touched. Depending on the intensity and number of noxious stimuli, the sensitisation lasts for minutes, hours, days or weeks; in other words, it takes that long for the animal to become habituated. Seen at a *psychological level*, the sensitised animal is said to *learn about* and *remember* the *experience* caused by the noxious stimulus; and the learning and memory are overtly exhibited by its responses to the tactile, non-noxious stimuli and its resistance to habituation. (See Kandel & Tauc 1965, Pinsker *et al.* 1973 for the original research. Some of the many reviews and text-book treatments are Hawkins *et al.* 1993; Kandel 1991; and Dudai 1989, Chapter 4.)

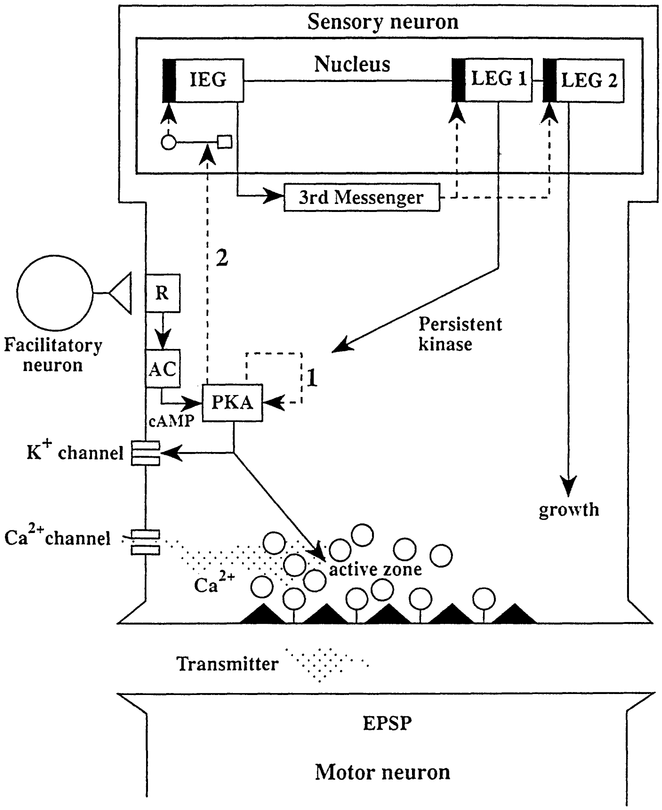
The goal of a neural and molecular analysis of the phenomenon of sensitisation is to find what internal mechanisms underlie the behavioural data, and thus what internal states can be identified as the cognitive mechanisms of learning and memory. The basic underlying anatomy includes several sensory, motor, and facilitatory neurons (as well as inter-neurons, which we need not worry about); in a schematic model, we shall consider one neuron of each kind, as illustrated in Figure 1 (page 172). The sensory neuron leads from the siphon and makes an excitatory synapse on the motor neuron, which in turn synapses on the muscle cells controlling movements of the gill; the facilitatory neuron leads from the head or tail and synapses on the terminal of the sensory neuron.

When the siphon is touched under normal circumstances — *i.e.*, prior to habituation or sensitisation — the sensory neuron fires and evokes an *excitatory post-synaptic potential* (EPSP) in the motor neuron; this causes the motor neuron to fire, the muscle cells to contract and the gill to withdraw. With habituation, the sensory-motor synapse *weakens*; that is, an action potential of a certain magnitude in the sensory neuron evokes a smaller EPSP in the motor neuron. With sensitisation, caused by noxious stimuli to the head or tail followed by non-noxious stimuli to the siphon, the synapse *strengthens*, in that the EPSP becomes greater; in turn, the EPSP causes a greater frequency of firing in the motor neuron, a more vigorous muscle contraction and hence withdrawal of the gill. The strengthening of the sensory-motor connection is called "synaptic facilitation" (for a reason to become clear presently).

In summary, there is a neural circuit involving the sensory, the motor, and the facilitatory neurons, such that the sensory-motor connection weakens with repeated firing of the sensory neuron, and strengthens when the firing of the sensory neuron is preceded by a firing of the facilitatory neuron. The issue of accounting for the behavioural phenomenon thus reduces to that of accounting for the variations in strength, or *plasticity*, of the sensory-motor synapse.

The strength of the synapse depends on the quantity of neurotransmitter released from the sensory-neuron terminal when that neuron fires: the more transmitter released, the greater the EPSP produced in the motor neuron; the less transmitter, the smaller the EPSP. The quantity of neurotransmitter released is modulated by the quantity of calcium ions (Ca^{2+}) which enter the terminal, *via* voltage-gated Ca^{2+} channels, when the sensory neuron fires. Under normal circumstances of transmission, the action potentials arriving at the terminal open both the Ca^{2+} channels, causing the release of neurotransmitter, and voltage-gated potassium (K^+) channels, the

Figure 1: *Mechanism underlying sensitisation in* Aplysia. The activated receptor (R) stimulates the enzyme adenylyl cyclase (AC) to synthesise the second messenger (cAMP); cAMP activates protein kinase *A* (PKA), which closes the K^+ channel and also helps to mobilise synaptic vesicles at the active zone. The closure prolongs the opening of the Ca^{2+} channel, and hence more calcium enters into the terminal and more transmitter is released. Maintenance is achieved, in the short term, by the conversion of PKA into a constitutively active form (pathway 1). In the longer term (pathway 2), PKA acts on nuclear substrate proteins (a box linked to a circle), which activate immediate-early genes (IEGs). The IEGs synthesise third messengers regulating the expression of late-effector genes (LEGs 1 and 2). The protein products of the LEGs then maintain the kinase in an active state, and effect the growth of new active zones and synaptic contacts.



function of which is to allow K^+ ions to enter the terminal, thereby restoring normal membrane potential, closing the Ca^{2+} channels and ending the release of transmitter.

When the facilitatory neuron is active, in response to noxious stimulation, the neurotransmitter released from its terminal — the so-called "first messenger" — binds to a receptor protein (R) in the membrane of the sensory-neuron terminal. The receptor then activates several *second--messenger pathways*. For our schematic purposes, we shall consider only one such pathway. The receptor is linked to a membrane-bound enzyme, adenylyl cyclase (AC), the function of which is to synthesise cyclic AMP (cAMP), which is regarded as the second messenger. The cyclic AMP then activates protein kinase *A* (PKA), which phosphorylates the K^+ channel, causing it to close. Since this channel, when open, works to restore normal membrane potential, the effect of closing it is that — when the sensory neuron fires in response to the non-noxious stimulation — the action potential lasts longer in the neuron's terminal; so the Ca^{2+} channel is open longer, more calcium enters the terminal, and more transmitter is released at the sensory-motor synapse. This causes a greater EPSP in the motor neuron, and hence a greater frequency of firing, more vigorous muscle contraction and withdrawal of the gill. (See Shuster *et al.* 1985; Siegelbaum *et al.* 1982; Bailey *et al.* 1983.)

Notice that this account of the cognitive mechanism underlying sensitisation is thoroughly *behaviouristic*. Experiencing is regarded as a firing of the sensory neuron; learning as the acquisition of a *behavioural disposition* physically implemented as a strengthening of the sensory-motor synapse; and memory, as we shall see anon, is regarded as the maintenance of the disposition. There is no room here for such conceptual and propositional states as believing that so-and-so is the case, desiring that so-and-so be the case, expecting that so-and-so will be the case, *etc.*, unless one allows — as some connectionists do — that these *prima-facie* complex representational states are to be identified with the synaptic states themselves, or with patterns of such states, or with emergent properties of such patterns of states.

In order to understand the possible *role of genes in cognition*, the important problem is that of the *maintenance* of the behavioural disposition, which is — so far as connectionists are concerned — the problem of memory in *Aplysia*. The duration of the disposition depends on the duration of the phosphorylation of the K^+ channel. However, without a continued reinforcement from the cAMP-dependent protein kinase, the phosphorylation lasts only from a few minutes to an hour. Research into the retention of synaptic facilitation indicates that there are two kinds of mechanism whereby the duration of the disposition is prolonged.

Firstly, there are several molecular mechanisms which prolong the activity of the kinase beyond the time-course of the elevated levels of cAMP; the kinase then maintains the modification of the K^+ channel and hence the behavioural disposition (Figure 1, pathway 1). These mechanisms are independent of the synthesis of new proteins, and can account for retention lasting from several hours to one day (Greenberg *et al.* 1987; Schwartz & Greenberg 1987).

Secondly, the activated kinase phosphorylates, in addition to the K^+ channel, a nuclear substrate protein (pathway 2) which acts as a regulator of the transcription of *immediate-early genes* (IEGs). In turn, the protein products of the IEGs are regulators, regarded as third messengers in the pathway from extracellular transmitters to intracellular genetic responses, that determine the patterns of expression of at least two classes of *structural*, or *late-effector* genes (LEGs 1 and 2). The one class of genes encode proteins which maintain the activity of the kinase, and therefore the phosphorylation of the K^+ channel. The other encode proteins which are transported to the terminal of the sensory neuron, where they effect two kinds of morphological change: an increase in the number of active zones — that is, zones which are capable of releasing neurotransmitter — in existing synapses; and a growth of new pre-synaptic terminals, matched by a growth of new dendrites in the motor neuron. (Montarolo *et al.* 1986; Goelet *et al.* 1986; Bailey & Kandel 1993; Kandel 1989; Montminy & Bilezikjian 1987).

The facilitated state of the sensory-motor connection is thus underlaid by a complex pattern of expression of the late-effector genes. This mechanism is akin to that involved in the ontogenetic development of an organism, where permanent rather than transient changes in patterns of gene expression underlie permanent changes in cellular form and function. As such, the mechanism could maintain the facilitated state, and hence the behavioural disposition, for long periods of time, even up to the life-time of the organism. (For the link between ontogenetic development and memory, see Goelet *et al.* 1986; Kandel 1989.)

Sensitisation is a *non-associative* form of learning: the noxious stimulus produces a facilitation in the sensory-motor synapse, such that a mild stimulus to the siphon, applied at any time during the period of retention of the facilitated state, elicits an enhanced behavioural response. The molecular mechanism of sensitisation can be elaborated to subserve more complex, *associative* learning, a form of classical Pavlovian conditioning. Here the temporal sequence of the stimuli is reversed. The mild stimulus — called "conditioned stimulus" — precedes the noxious unconditioned stimulus by about 0.5 seconds. After such training, the mild conditioned stimulus, applied by itself, elicits an enhanced response. This form of learning is more effective than simple sensitisation, in that the response is stronger and longer lasting.

The key molecular processes underlying conditioning are as follows. Action potentials arriving at the sensory-neuron terminal as a result of the mild conditioned stimulus occasion an influx of Ca^{2+} into the terminal. The calcium binds to calmodulin to form a Ca^{2+}/calmodulin complex, which in turn binds to the enzyme adenylyl cyclase. The enzyme undergoes a conformational change, such that when it is activated by the incoming unconditioned stimulus — *via* the facilitatory neuron — it synthesises more cyclic AMP, and thus causes a stronger activation in a greater number of PKA molecules. The rest of the sequence is like that involved in sensitisation. The PKA closes the K^+ channels, so allowing of a greater influx of Ca^{2+}, a greater release of transmitter, and stronger EPSP in the motor neuron. In addition, the PKA activates LEGs 1 and 2, which in turn maintain the activity of the kinase and cause new synaptic growth. (See Kandel 1991, Hawkins *et al.* 1993.)

To summarise: both associative and non-associative learning in *Aplysia* require, in the longer term, the formation and maintenance of underlying patterns of gene expression. Hence it is not far to the thought that since discrete synaptic states — which underlie discrete patterns of behaviour — are underlaid by discrete patterns of gene expression, we could attempt to identify cognitive states at the genetic level. However, the processes activating gene expression are rather slow, so that cognition would appear to lag behind behaviour; also, it may not yet be clear why the genetic level should be preferred to the synaptic level. I will consider these issues in Section 8.4; but prior to that, it will be useful and instructive to review some aspects of cognitive implementation in higher animals and our own species.

8.3 Learning, Memory, and Association in Vertebrates

There is a variety of studies which indicate that the hippocampus, a subregion of the temporal lobe on each side of the brain, is the organ where new declarative, propositional cognitive states are *formed*; and that the temporal lobe is the region where long-term declarative memories are *stored*. As regards the *storage* of long-term memories, there are the classical studies of localisation of cognitive functions carried out by Wilder Penfield in the course of neurosurgery on the temporal lobes of patients suffering from epileptic seizures. To avoid damage to functionally significant areas, Penfield used small currents from electrodes to stimulate minute portions (in effect, single neurons) of the exposed lobes. The patients — under local anæsthesia but fully conscious — often reported having detailed recollections of specific experiences of events in their past. For example, a patient might recall listening to music at a concert hall, or watching a baseball game many years ago, *etc.* (see, for example, Penfield & Roberts 1959). As regards the *formation* of new memories, there are the cases of patients whose hippocampus was removed on both sides of the brain as a remedy for epilepsy. These patients lose the ability to form new long-term declarative memories, but hold intact their store of memories acquired prior to the operation; they also retain a normally functioning short-term, or working memory — which is thought to be located in the pre-frontal lobes of the cortex — but cannot transform newly acquired information into long-term storage (Milner 1966; Squire 1992).

The picture that is beginning to emerge is that new cognitive representations are formed in the working memory of the pre-frontal cortex as a result of activity in the sensory cortices. From working memory, the sensory information is relayed to the hippocampus, which stores and processes it up to several weeks or months, and gradually transfers it to other cortical regions for long-term or permanent storage. The sites of permanent storage are again connected to the working memory, so that old information may be retrieved whilst new is being acquired (see Goldman-Rakic 1987 for a detailed treatment).

We shall now turn to consider the issue of memory formation and retention in the hippocampus. The hippocampus comprises several neural pathways connected by synaptic junctions which are thought to play a key role in the formation and processing of new long-term memories. A striking property of the synapses is that they are capable of *long-term potentiation* (LTP). LTP is produced by stimulating pre-synaptic neurons with a brief, high-frequency train of electric shocks, called "tetanic stimulation". The strength of a tetanised synapse increases immediately about fivefold, but settles down within minutes to a level of about twice the original. This initial increase is called "post-tetanic potentiation" (PTP), and LTP is defined as any increase lasting longer than PTP (Bliss & Lømo 1973; Bliss & Lynch 1988; Madison *et al.* 1991). (In what follows, we shall confine our attention to the form of LTP which occurs in the dentate gyrus and CA_1 region of the hippocampus; for other forms of LTP and the relevant distinctions, see Johnston *et al.* 1992.)

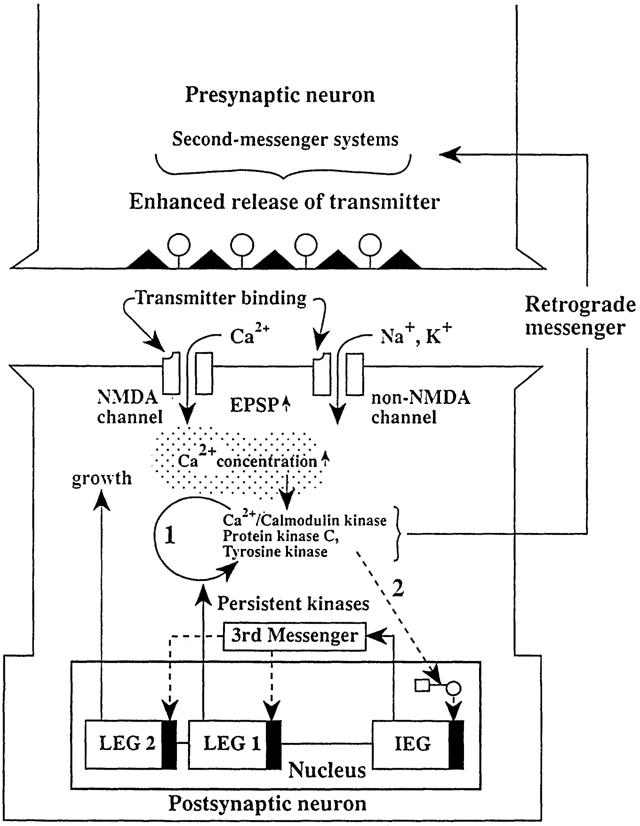
Notice that LTP is similar to, but not quite the same as, synaptic facilitation in *Aplysia*. One of the differences is that LTP is produced by tetanic use of a synapse; in *Aplysia*, use of the sensory-motor connection brings about habituation, or weakening, whereas facilitation is caused by the activity of the facilitatory neuron. However, both LTP and facilitation are cases of synaptic plasticity, and both are regarded, in connectionist psychology, as a neural basis of learning and memory. The working hypothesis is that *learning*, or the acquisition of a cognitive state, consists in the *induction* of (a pattern of) alterations in the strength of certain synaptic connections; and *memory* in the *maintenance* of those alterations.

The *induction* of LTP is found to depend on the functioning of two classes of post-synaptic receptor proteins: the NMDA receptors, so called after N-methyl-D-aspartate which selectively activates them, and non-NMDA receptors. We shall consider only one receptor of each class (see Figure 2, page 178). Both receptors are channel-linked; that is, they are coupled to a transmembrane ion channel. The non-NMDA channel is ligand-gated, in that it opens in response to the binding of an appropriate ligand, which in our case is the neurotransmitter. When open, the non-NMDA channel is permeable by sodium (Na^+) and potassium (K^+) cations, but not by calcium (Ca^{2+}) and other divalent cations. In contrast, the NMDA channel is both ligand- and voltage-gated. The effect of this is that the binding of the neurotransmitter is not sufficient to open the channel; to open the voltage gate, it is also necessary that the post-synaptic membrane be strongly depolarised. When the channel is open, in response to transmitter binding and depolarisation, it carries not only Na^+ and K^+, but also Ca^{2+}. The induction of LTP occurs only when a critical level of Ca^{2+} current flows into the post-synaptic neuron through the NMDA channel.

It is because the NMDA channel opens only when two independent conditions are satisfied (*viz*., post-synaptic depolarisation and pre-synaptic release of transmitter) that LTP, of the form we are considering, is regarded as a mechanism with potentially *associative* properties. For these conditions amount simply to the requirement of *coincident activity* in the pre-synaptic and post-synaptic cells; and, in turn, this can be achieved in either of two ways. Firstly, persistent tetanic firing of the pre-synaptic neuron will produce sufficient coincident depolarisation in the post-synaptic neuron, as well as release transmitter. Secondly, the post-synaptic neuron may receive depolarising inputs from another, associated source, coincidently with even a brief input from the pre-synaptic neuron, which on its own would not be strong enough to produce sufficient depolarisation. The second way clearly lends itself to associative learning.

We are now in a position to set out the basic molecular events leading from tetanic stimulation to induction of LTP (see Figure 2). The tetanic stimulus is a rapid series of individual action potentials, each of which releases an amount of transmitter from the pre-synaptic terminal. The transmitter — the first messenger — diffuses across the synaptic cleft and binds to both the NMDA and non-NMDA receptors. The non-NMDA receptor channel opens, and a current of Na^+ and K^+ flows into the post-synaptic cell, thereby depolarising the membrane and producing an EPSP. When the EPSP reaches a certain level, the voltage gate of the NMDA channel opens, and a current of Ca^{2+}, in addition to the monovalent ions, flows into the cell. The calcium ions then activate, either directly or *via* a second messenger, at least three kinds of protein kinase: Ca^{2+}/calmodulin kinase, kinase *C*, and tyrosine kinase. The kinases in turn cause the synthesis of a retrograde messenger, thought to be nitric oxide. The messenger diffuses across the synaptic cleft to the pre-synaptic terminal, where it acts to facilitate — perhaps *via* further second-messenger systems — the release of transmitter. The initial stages of LTP, although mediated by a post-synaptic mechanism, thus consist in a modulation of transmitter

Figure 2: *LTP in the hippocampus.* Neurotransmitter binds to both the NMDA and the non-NMDA receptors. The non-NMDA receptor channel opens, and a current of Na$^+$ and K$^+$ flows into the cell, so depolarising the membrane. The voltage-gate of the NMDA channel opens, and a current of Ca^{2+} enters the cell. The elevated concentration of Ca^{2+} activates Ca^{2+}/calmodulin kinase, protein kinase *C*, and tyrosine kinase, which cause the synthesis of a retrograde messenger. The messenger then diffuses to the pre-synaptic terminal, where it facilitates the release of transmitter. In the short term, maintenance is achieved by the conversion of the kinases into constitutively active forms (pathway 1). For long-term retention (pathway 2), the kinases act on nuclear substrate proteins (a box linked to a circle) which regulate the expression of immediate-early genes (IEGs); in turn, the IEGs synthesise third-messenger proteins which regulate the expression of late-effector genes (LEGs 1 and 2). The LEGs then maintain the activity of the kinases and effect new synaptic growth.



release from the pre-synaptic neuron. (Haley *et al.* 1992; Schuman & Madison 1991. For reviews see Hawkins *et al.* 1993; Madison *et al.* 1991. Textbook treatments are Kandel 1991; Dudai 1989, Chapter 6.)

Following induction, the pre-synaptic modification can account for a short-term retention of the altered synaptic state, lasting from several minutes to an hour. This is called "decremental LTP", or "short-term potentiation" (McNaughton 1982). Research into the *maintenance* of LTP shows that there are two kinds of mechanism of long-term retention, similar to those involved in long-term facilitation of the sensory-motor connection in *Aplysia.*

Firstly, the kinases activated by the influx of calcium may be converted, provided the incoming stimulus is sufficiently strong and enduring, into constitutively active forms in which they remain operative even in the absence of the stimulus, so maintaining the synthesis of the retrograde messenger, and thereby prolonging the time-course of the pre-synaptic facilitation (Figure 2, pathway 1). This mechanism can account for retention lasting several hours. (Schwartz & Greenberg 1987; Malinow *et al.* 1988; Klann & Sweatt 1990.)

Secondly, the kinases act on nuclear substrate proteins that regulate the expression of IEGs (pathway 2). The protein products of the IEGs — the third messengers — then determine the patterns of expression of the late-effector genes (LEGs 1 and 2). These in turn undergo transcription and translation, and their protein products are transported back to the stimulated synapse to maintain the potentiated state and to initiate new synaptic growth. As in *Aplysia*, the function of maintenance in the long term is thus carried by altered patterns of expression of the underlying effector genes, in a way similar to the maintenance of cell differentiation. (Armstrong & Montminy 1993; Bailey & Kandel 1993; Cole *et al.* 1989; Frey *et al.* 1988; Madison *et al.* 1991.)

The fact that long-lasting changes in synaptic states, which are regarded as mechanisms of learning and memory, are underlaid by long-lasting changes in genetic states, indicates that memory and other cognitive states might be more correctly identified as genetic rather than synaptic states. My plan in the rest of the chapter is not to add to the evidence, which is something for molecular biologists to do, but rather to use the Classical Theory of Mind to outline how such psychological states as beliefs and desires could be implemented in the brain at the genetic level of analysis, and to show why the genetic level is preferable to the synaptic level as the appropriate level for cognitive individuation.

8.4 The Genetic Code, the Mental Code

There is a traditional account of mind which suggests itself in the light of common-sense, introspective psychology; namely, that the mind is a system of operations, identifiable with such introspective functions as believing, desiring, perceiving, and so forth, on the sentences of a language-like representational code. This picture is implicit in most classical philosophy of mind. It was — for the first time, perhaps — *explicitly* formulated in Locke's *Essay*, and it began to be undermined to some extent in Hume's *Treatise*. But the first concerted attack on the classical theory came early in this century with behaviourism and behavioural holism in psychology and linguistics, and, above all, connectionism in neuroscience. I have already described the classical position in some detail in Chapter 7, though a full account will have to wait till Chapter 9; here I will begin by sketching only so many of its basic features as will be needed for the sake of this chapter, which is to show how the mind so construed could be implemented at the level of neuronal genes.

8.4.1 The Classical Theory of Mind.

Although the classical theories — of Descartes and Locke, among others — differ in many important details, at least the following they may be said to hold, more or less implicitly, in common concerning the constitution and functioning of the mind:

(α) there is a finite basis of semantically simple mental symbols; that is, of *simple terms* of the mind's representational code;

(β) there is a generative mechanism for the production of infinitely many *complex terms* from the basis of simples;

(γ) there is a generative mechanism for the production of infinitely many *sentences* from the simple and complex terms;

(δ) there is a basis of *simple operations* on *tokens* of the sentences, or *propositions*;

(ϵ) there is a generative mechanism for the production of *complex operations* from the basis of simple ones.

The operations, whether simple or complex, are then identified with such introspective functions as believing, desiring, and so forth. A particular mental state, such as believing that bats are birds, is identified with an instance of the belief-operation on a token of the mental sentence that means that bats are birds. The processes of thought, deliberation, reasoning, willing, *etc.*, are identified with causal sequences of instances of such operations.

The disagreement among classical theorists concerned, *inter alia*, the size and character of the basis of semantically simple terms, and the relationship between simple terms and experience. Contemporary classicists also argue about the issue of defining complex terms from the basis of simples (Fodor 1981), defining sentential symbols from the basis of terms (Chomsky 1986), and defining complex operations from simple ones (Searle 1983). Our present concern is the issue of the biological implementation of CTM, as expressed in clauses (α)–(ϵ). There are two aspects to this problem: firstly, an *ontological aspect*, regarding the kind of *entity* that could implement mental symbols, and operations on mental symbols, in the brain; secondly, a *psychological aspect*, regarding the kind of *cognitive architecture* that could implement such a system of symbols and operations. We shall start with the ontological aspect.

8.4.2 A conjecture on the nature of mental symbols.

It is remarkable that genes were originally posited, and are still typically regarded, as symbols of a sort: as *information-containing entities*, supposed to underlie heritable phenotypic properties and ontogenetic development of organisms. But it was not until the 1940's that they were recognised as molecules of deoxyribonucleic acid (DNA). DNA consists of two helical chains, each composed of four nucleotide units: adenine (*A*), thymine (*T*), guanine (*G*), and cytosine (*C*). The chains are held together by complementary pairing of the units: *A* pairs with *T*, and *G* with *C*. In the 1950's, it was proposed that DNA carries *biological information* encoded by sequences of *A, T, G*, and *C*. Later it was found that each of the twenty common amino acids which proteins are built of is encoded by a sequence of three nucleotides (with most encoded by more than one such triplet). The nucleotides were considered as *letters* of the alphabet of a *genetic code*, and the triplets — called "codons" — as basic *terms* of the code. Sequences of codons specifying entire proteins were sometimes considered as *sentences* of the code. (Yčas 1969; Crick 1963; Watson *et al.* 1987.)

It is not necessary to regard codons as *terms* of the genetic code, or protein-coding sequences of codons as *sentences* of the code. But it is important that such a biological system of representation could well *function as* a system of cognitive representation. In particular, a semantically simple term of the mental code could be construed as a certain pattern of expression of representations of the genetic code; in effect, as a sequence of the units *A, T, G*, and *C*. A complex term could be construed as a complex of such sequences, generated by a mechanism itself implemented — perhaps as a system of regulatory genes — in the genetic code. Similarly, sentences of the mental code could be construed as certain further structures of representations of the genetic code; and cognitive operations — such as believing, desiring, and so forth — could be construed as genetic operations on tokens of the sentences.

8.4.3 Cognitive architecture.

In the remaining parts of this chapter, I will try to make such a construal plausible, insofar as the speculative character of the proposal allows. My contention will be centred on the second aspect of the question of biological implementation, regarding the kind of cognitive architecture that would comport with CTM. I will reject the *classical computational architecture* (CCA), which is usually taken to be *required* by CTM, and instead give an answer motivated by the genetic conjecture. The resulting account, which I will call "the genetic theory of mind" (GTM), will be seen to have, among others, three commendable properties:

(a) it accommodates common-sense psychology and allows of a natural account of rational processes;

(b) it satisfies the connectionist insight that most mental processes are distributed over several functional areas of the brain, and many processes run in parallel;

(c) it accords with such data concerning the biological basis of cognition as we have reviewed in Sections 8.2 and 8.3.

Let us now turn to see what cognitive architecture will suit CTM. The classical theory of mind is often believed to *require*, or even be identical with, the classical computational architecture (for the former option, see Fodor 1987, pp. 16–19, 139). But in fact CTM and CCA are distinct, independent doctrines: CCA does not require CTM, and CTM does not require CCA.

It is clear that CCA is independent of CTM; for the symbols and operations of CCA *may*, but *need not*, be interpreted as the posits of CTM (that is, as the posits mentioned in clauses (α)–(ϵ)). It is more difficult to show that CTM does not require CCA. In order to do that, one has to offer an alternative to CCA, agreeing with CTM. I will turn to elaborate such an alternative anon. But firstly, it will be helpful to recall and refresh our motivation to seek a new cognitive architecture for CTM. The motivation comes from the connectionists' most persuasive argument (CMPA), which says that the classical paradigm cannot be correct, for two reasons. Firstly, there are no entities and processes in the brain identifiable with tokens of the terms and sentences of a representational mental code, and operations on sentence-tokens. Secondly, there is no faculty in the brain identifiable with a central processor in which mental symbols, if there were any, could be displayed and operated on (see Section 8.1 for CMPA). (The working memory of the pre-frontal lobes is sometimes considered as serving a similar role, but there is no evidence that it functions as a *central processor* of the CCA-kind.)

The appeal of this argument stems, in its second part, from the mistaken assumption that the classical paradigm *amounts to* CCA, or — where CTM is considered — the assumption that CTM *requires* CCA. For, granted either of these assumptions, the second part of CMPA is very cogent: it is unlikely that there is a central processor of the CCA-kind in the brain. Therefore, if the classical paradigm is to be maintained, CTM will have to be given a new cognitive architecture. I will set out my proposal for CTM's cognitive architecture by means of a simple schema designed to accord with, and satisfy the needs of, the genetic conjecture; and I will divide the proposal into two components, to be regarded metaphorically as concerning the *vertical* and the *horizontal* dimensions of the architecture.

8.4.3.1 The vertical dimension.

The basic notion of the schema is that of a *psychic cell.* (The term "psychic cell" is borrowed from Ramón y Cajal; we shall come to the historical aspects of the proposal in Section 8.5.) A psychic cell is a neural cell which:

- *(i)* is specialised for the function of mental representation;
- *(ii)* forms, stores, and processes representations as patterns of expression of its genetic material.

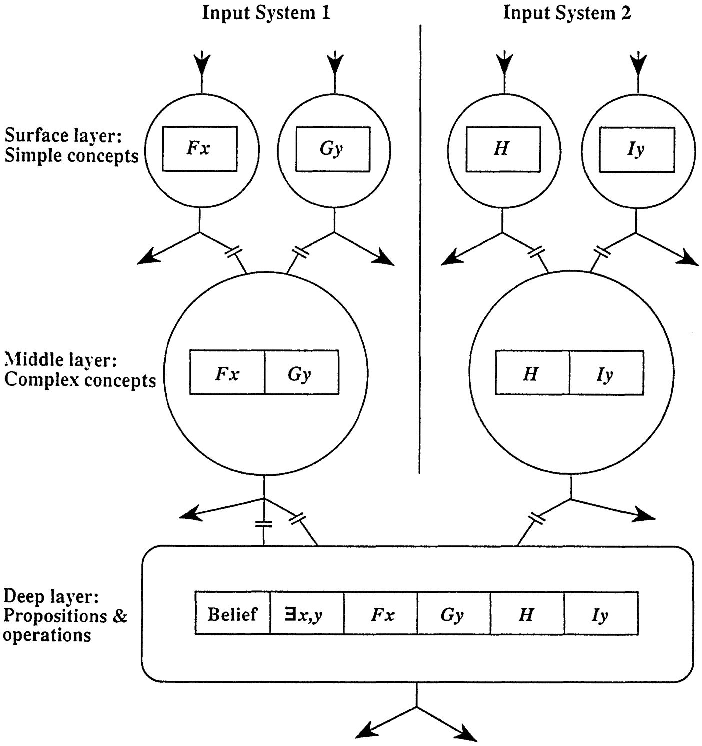
According to the schema, illustrated in Figure 3 (p. 184), there are three classes of psychic cells, corresponding to clauses (α)–(γ) of the classical theory:

- I. Input cells, each of which carries a *single, semantically simple, empirical term.* We may think of the input cells as arranged in a *surface layer* of the mind's representational system.
- II. Middle-layer cells, each carrying some part of the empirical basis of semantically simple terms, together with a generative mechanism for the production of *complex terms.*
- III. Deep-layer cells, each carrying the entire basis of semantically simple terms, together with a mechanism for the production of *sentences* from terms; and the basis of *simple operations* on tokens of sentences (propositions), together with a mechanism for the production of *complex operations.*

The input cells of the surface layer are ordered and connected to the middle-layer cells in such a way that, when several of them are activated in response to external stimulation, each will pass some signal to one or more middle-layer cells, with the effect of *transferring a token of the same symbol-type* from the input cell to the middle-layer cell(s). Later I will suggest that, in the initial stages of symbol-formation, the signal passed is chemical and slow, whereas in late stages — or in the recall of a familiar symbol — it is voltage-dependent and fast; for the moment, the nature and speed of the signal need not concern us. Similarly, when several of the middle-layer cells are activated, each will transfer the *complex symbol* formed within it as a result of the surface inputs, to one or more deep-layer, proposition-forming cells; and the transfer will take place by sending a set of signals from the middle-layer cell to the deep-layer cell(s), causing the latter to express tokens of the same symbol-types. The deep-layer cell(s) will further process the inputs, forming *propositions* and *complexes* of propositions.

There are two points about the vertical dimension I would like to make clear. The first concerns the notion of transferring tokens of the same symbol-type between psychic cells: from a cell carrying a single semantically simple symbol to one carrying a complex term, to one carrying a sentential symbol. The transfer is assumed to occur not by moving the tokens between the cells, but by means of genetic expression: *i.e.*, when an input cell is activated and thus tokens the simple symbol it carries as a pattern of gene expression, it passes a signal to each of the cells in the middle layer to which it is connected, and this causes the middle-layer cells to token a type-

Figure 3: *Psychic-cell architecture, the vertical dimension.* The illustration involves three sets of psychic cells, arranged in three layers. The surface- and middle-layer cells are further arranged in two input systems, which can be thought of as visual, somatic, auditory, *etc.* The input cells of the surface layer each encode a single semantically simple term, *F, G, H,* or *I.* Variables indicate which cells are active, and thus which terms are tokened; so the cell carrying *H* is not active, and *H* is not tokened. When a surface-layer cell is activated, it passes a signal to the connected middle-layer cells, causing each to token a type-identical term by expressing a type-identical gene. Similarly, an active middle-layer cell causes the deep-layer cells to which it is connected to token terms of the same type. The deep-layer cell is pictured as further processing the inputs, adding the quantifier closure over x, y, and tokening the proposition $(\exists x, y)(Fx \wedge Gy \wedge Iy)$ under the operation of believing. The generative mechanisms of the middle and deep-layer cells are not represented in the illustration, nor are complexes of propositions and propositional operations.



identical symbol by expressing a type-identical gene. Likewise, when a middle-layer cell is activated and thus tokens the complex term it carries as a complex pattern of gene expression, it passes — perhaps by way of a discrete synaptic connection for each simple constituent of the term — a complex pattern of signals to each of the cells in the deep layer to which it is connected, and this causes the deep-layer cells to token the same complex term by expressing the same complex gene. (The notion of transferring tokens of the same symbol-type is reminiscent of Hume's suggestion that mental representations stored in memory, which he called "ideas", are faint resemblances or *copies* of surface representations, which he called "impressions", and which he regarded as directly excitable by sensory inputs; see Hume 1975, Section II; 1978, Book I, Part I, Section I. Although my own account is by no means Humean, I think Hume did succeed in making explicit the notion that the mind comprises *several layers* of mental symbols, with at least some of the symbols in the deep layers being *copies* of those in the surface layers; this important insight was only implicitly present in Locke and Descartes.)

The second point concerns the formation and processing of *complexes of propositions* within a *deep-layer cell.* Such a cell is assumed to form — by its generative mechanism on simple and complex concepts (clause (γ) of CTM) — not only single but complexes of propositions; and the processing of the complexes is assumed to involve not single but complexes of cognitive and emotive propositional operations ((δ)–(ϵ) of CTM). Accordingly, the patterns of genetic expression which implement the system of generative mechanisms, propositions, and propositional operations (clauses (γ)–(ϵ) of CTM) in the deep-layer cell, are not to be thought of as static and immutable, but rather as comprising a *dynamic* genetic system capable of changes in response to further learning stimulations, and capable — once formed and stored in the cell — of internal processes which can be identified with cognitive processes. I will suggest later that simple *rational* (as opposed to *associative*) processes occur in such dynamic genetic systems within a deep-layer cell, whereas more complex rational processes occur in systems of such cells.

8.4.3.2 The horizontal dimension.

The horizontal dimension of the schema will be inspired by studies of localisation of cognitive functions in the brain, some of which I have briefly reviewed in Section 8.3. I mentioned there four areas of the cortex, each fulfilling a broadly specified function in the formation, storage, and processing of cognitive states:

(1) the temporal lobe, where long-term memories appear to be *stored*;

(2) the hippocampus, which is the long-term *memory former*;

(3) the sensory cortices, which may be viewed collectively as the *sensorium* of the mind, receiving external or internal inputs, and forming sensory-level mental representations;

(4) the pre-frontal lobes, which implement the mind's *working memory*, receiving information from the sensory cortices for short-term processing, relaying new information to the memory former, and reactivating old information in the long-term store.

Such a quadripartite division belies the intricacy of the brain's actual cognitive architecture (*cf.* Goldman-Rakic 1987), but it will be adequate for our schema. Our chief concern in setting out the horizontal dimension of the schema is to distinguish the role of working memory from that of a central processor in the classical computational architecture, and thereby to guard against the assumption that CTM requires CCA. To this end, working memory will be viewed not as a central unit in which symbols are displayed and operated on, and in which all cognitive processes take place serially, but rather as having the following goals:

(4a) to *mediate* between, and *coordinate* the activity of, the other three faculties: *viz.*, sensorium, memory former and long-term storage;

(4b) to *orchestrate the activity of psychic cells* in the long-term storage, primarily in response to information it receives from the sensorium, but also in response to information received from the memory former and the long-term store.

Function (4b) calls for particular attention. It determines which psychic cells in the long-term store become activated, and so which cognitive-genetic processes occur. Thus it ensures that the mind responds with *appropriate* cognitive processes to incoming external or internal stimuli, processes which take place within the psychic cells in the long-term store. In other words, working memory does not recall symbols *from* long-term storage, to be displayed and operated on in it as in a central processor, but — conversely — it activates appropriate psychic cells *in* long-term storage, and hence appropriate cognitive processes stored in the cells. This notion will be further clarified in Sections 8.4.4–8.4.6.

In order to integrate the horizontal and vertical dimensions of the architecture, I will suppose that the three-layer system of psychic cells structures each of the four faculties, allowing that the system may be flexible and serve distinct purposes in distinct faculties. For example, it is plausible that the deep-layer cells in the long-term storage form very complex and stable cognitive-genetic modules, whereas in the sensorium they form less complex, shorter-lived representational patterns. However, I need not specify what the distinct purposes are. Similarly, it is not at present necessary to tell just how the four faculties communicate information one to another: *i.e.*, how the layers of psychic cells in one faculty are linked to those in another;

and so on. Such issues I will leave for a more detailed formulation of the schema.

The vertical and horizontal dimensions together comprise what I will regard as the *psychic-cell cognitive architecture* (PCA); and the combination of PCA, CTM, and the genetic ontology of mental symbols and operations I will regard as the *genetic theory of mind* (GTM). GTM is clearly a version of the classical paradigm (with PCA replacing CCA), though it does accommodate, as we shall see anon, the major insights of connectionism. I will now turn to indicate, briefly, how GTM accounts for such cognitive acts as concept-acquisition, memory, recall, associative processes, and rational processes.

8.4.4 Concept-acquisition and memory.

I mentioned earlier that a key aspect of the vertical dimension of PCA is that of transferring tokens of the same symbol-type between psychic cells: from a cell carrying a single semantically simple symbol to one carrying a complex term, to one carrying sentential symbols. Such a transfer requires that there be discrete links between type-identical symbols within connected cells. Since these links must run *via* synaptic connections, the best candidates would seem to be the second- and third-messenger pathways that link synapses with underlying effector genes, and the protein-transport pathways that target proteins to make new synaptic contacts and to modify existing synapses. However, the messenger pathways are very slow; taken together, they require from minutes to hours to execute, and the transport of newly made proteins is yet slower (see, *e.g.*, Morgan & Curran 1991; Alberts *et al.* 1989, Chapter 19). This shows that although the chemical pathways could function in the *formation* of cognitive representations — *i.e.*, in concept-learning — they could not function in relating *acquired* cognitive states to behaviour: that is, in recalling appropriate states in response to experiential stimuli, and mediating behavioural outputs.

According to GTM, the learning of new concepts is a process of the formation of new patterns of gene expression, and memory is the retention of the acquired patterns. The formative processes are slow, and may be subserved by the chemical pathways that connect synapses with underlying patterns of gene expression. Once such genetic-cognitive patterns are acquired, they may be retained up to the life-time of an organism by means of mechanisms similar to those involved in the maintenance of altered patterns of gene expression in cell differentiation.

8.4.5 Recall and simple associations.

One way acquired cognitive states, *qua* genetic states, could be related to experience and behaviour is by *voltage-dependent processes*. When a cell receives depolarising inputs, the evoked EPSPs summate to yield the *grand post-synaptic potential* (GPSP), causing the cell to fire if a critical level of GPSP is reached. Given that there must a way of activating cognitive states which is not dependent on the slow, chemical pathways, it is plausible to suppose that cognitive responses are triggered by some conversion, in the cell nucleus, of the GPSP to processes that activate the cognitive-genetic states. The critical level of GPSP needed to elicit cognitive responses could be similar to that required for cell firing.

This supposition would also help to explain how a propositional representation stored in long-term memory might be *recalled* in consequence of a familiar occasioning experience. The proposition is assumed to be *formed* in the deep layer of the long-term memory store by transferring its component concepts from the middle layer; these, in turn, are formed by transferring their components from the surface-layer cells, and the simple concepts of the surface-layer cells are triggered directly by the occasioning stimuli. In this procedure, a pattern of synaptic *associations* is established between the input and the middle-layer cells, and between the latter and the deep-layer cell containing the proposition, which facilitates the spread of activation from the input cells so that it converges on the deep-layer cell, whenever similar occasioning stimulations reoccur after the learning is completed. So the cell which carries the proposition will be more likely to reach the critical level of depolarisation needed to activate the cognitive processes in its nucleus, when an instance of the occasioning state of affairs is presented.

8.4.6 Rational processes.

A key difference between GTM and standard versions of the classical paradigm, which accept CCA, is that the latter regard rational processes as occurring in a central processor, whereas GTM holds that such processes take place in individual deep-layer psychic cells, implemented as sequences of genetic operations on genetic symbols. The role of working memory, according to GTM, is not unlike that of a conductor in an orchestra: it is to determine which psychic cells are active, and hence which cognitive processes occur, in response to both incoming sensory stimulations and other occurrent thoughts.

I have already emphasised (8.4.3.1) that in order to appreciate the idea of rational processes as genetic processes, we should think of the acquired and stored patterns of genetic expression in a psychic cell not as fixed and immutable, but rather as comprising a complex and dynamic genetic system capable of alterations and processes within itself. More complex processes could be formed by the association of several psychic cells *via* synaptic junctions; and still more complex processes — more properly rational processes — could be orchestrated in a determinate manner by the working memory. Many processes could be under way in parallel, with most distributed over several functional areas of the brain. The storage capacity and information processing power of such a hypothetical mental system would be immense, even assuming, as we must, that only a fraction of the genetic material in a psychic cell is used for mental representation and processing, and only a fraction of neural cells are specialised to function as psychic cells.

It remains to point out that GTM has the desirable property of accommodating both the virtues of the classical paradigm and the insights of connectionism, as well as satisfying such constraints of cognitive biology as we have reviewed in Sections 8.2 and 8.3. The main virtues of the classical paradigm are that it comports well with common-sense psychological explanation, and that it allows of an account of rational processes as sequences of operations on tokens of mental sentences; and these are clearly preserved in GTM. It is also clear that GTM caters for the connectionist insight that cognitive processes in the brain are neither serial nor discretely localised, but massively parallel and distributed. Thus GTM could be seen to be a convergence of the classical paradigm and connectionism. Lastly, GTM accords with, and is responsive to, recent research into the biological basis of cognition, which implicates genetic expression in the formation, maintenance, and processing of mental states.

8.5 History, Histology, and the Molecular Level of Analysis

This section will review some of the history of proposals more or less akin to GTM. There are two periods of thought requiring a special attention: the first involves the early psycho-histology of Ramón y Cajal; the second, the attempts in the 1960's to identify mental states in the brain at the molecular level of analysis. Having surveyed these, I will finish with some reasons why the earlier proposals were not — in their time — successful, and some comparisons between GTM and the earlier positions.

Cajal's psycho-histological theory is set out at a number of places in his writings, but perhaps the most complete expositions are in Cajal (1990/1894: Chapter 3; 1988/1911: 479-87, 382-413). I will summarise, firstly, those features of his account which agree with and anticipate GTM; secondly, those features which make his account incompatible with GTM. There are four areas of agreement:

(i) Cajal takes it that the mind, as implemented in the brain, is a system of *psychic cells*, which are neural cells specialised for the function of mental representation (see Cajal (1990/1894: 47, 67-70)).

(ii) The psychic cells are arranged in several layers. Cajal (1990/1894) distinguishes four layers of psychic cells, arranged in four layers of the cortex: a superficial or molecular zone; a layer of small pyramidal cells; a layer of large pyramidal cells; and a layer of polymorph cells (pp. 39-52). In (1988/1911), Cajal recognises seven layers, which are those here mentioned, but with sub-layers noted in the second, third and

fourth layers (p. 383). (For a contemporary classification of cortical layers see, for example, Crick & Asanuma (1986: 341-2), who identify four layers: a superficial, an upper, a middle, and a deep layer.) I do not wish to map my three-layer schema, in any simple way, onto the cortical layers of Cajal's psychic cells. The reason is that the schema is designed primarily to satisfy the needs of CTM (clauses (α)-(γ)). But it is not implausible to think of my surface layer as, perhaps, Cajal's superficial zone, my middle layer as his layer of small pyramidal cells, *etc.* — especially considering Cajal's view that cognitive processing begins at the first layer and proceeds to the deep layers (1990/1894: 61), and considering that the size and complexity of Cajal's psychic cells increase from the surface to the deeper layers (*ibid.*: 48-50).

(iii) Cajal assumes the existence of cognitive representations (ideas, concepts), and he regards representations as carried by psychic cells in the details of their chemical composition — that is, at a molecular level of analysis — rather than at the level of their inter-connections (1990/1894: 67-8; 1988/1911: 479).

(iv) Historically, Cajal was not in a position to seek the identity of cognitive symbols in the patterns of genetic expression within psychic cells; but he linked the formation and maintenance of new representations — that is, learning and memory — to the ontogenetic development of psychic cells (1990/1894: 70-71; 1988/1911: 484-87). This is not to say that Cajal held the genetic hypothesis; but it does show that he considered whatever mechanisms underlie cell differentiation as necessary for the formation and maintenance of representations.

So much for the points of agreement. Contrary to GTM, Cajal was an *associationist*, and this position is not to be identified with either the classical paradigm or connectionism. He took each psychic cell to carry a single semantically simple symbol, or 'simple image of a sensory impression' (see 1990/1894: 71); and he took complex symbols and cognitive states to be formed by the association of several psychic cells. Cajal would have regarded his position as classicist; but this shows only the pervasive misunderstanding of the classical theory in the 19th and the present centuries (for which, I dare say, Hume is chiefly to be blamed). In contrast, and notably, Cajal did not subscribe to connectionism. In (1988/1911: 479-484), he debates three conjectures concerning the mechanisms of learning and memory: those of Duval and Tanzi (shrinking or lengthening of neuronal arborisations, and thus of the gaps at neuronal connections); and that of Lugaro (chemical alterations at connections, causing changes in the strength of transmission, with transmission over the gaps being chemical). All of these are early versions of connectionism, and Lugaro's view is very close to modern synaptic theory. Cajal's position differs in two respects. Firstly, he stresses the role of the growth and development of new neuronal connections as a result of 'cerebral gymnastics' (*i.e.*, learning); yet the connections function only to establish novel associations among ideas, not as sites of psychic acts (1988/1911: 484-87). Secondly, as mentioned earlier, the ideas are thought to be carried in the protoplasm of psychic cells by means of some as yet unknown molecular substrates.

After Cajal, associationism gave way to connectionism with the focus on synaptic transmission, and the last link with the classical theory — Cajal's ideas or representations in the inter-connected neurons — was abandoned in favour of behaviourism, or behavioural holism, or (in philosophy) emergentism of some sort. It should be recognised that this was a process of a simplification of the task at hand, of exorcising the ghost, so making the subject matter of cognitive neuroscience amenable to available means of research.

I will now turn to the second period of thought reminiscent of GTM, in which attempts were made to identify cognitive states at a molecular level of analysis. This began in the late 1950's and early 1960's with discoveries that RNA and protein synthesis are involved in learning and memory, and suggestions that either RNA or proteins could encode cognitive information, analogously to the encoding of genetic information in the structure of DNA (for a review, see Dunn (1976; 1986)). The studies were of two kinds: those attempting to find specific *correlations* of RNA or protein synthesis with learning and memory; and those showing that learning and memory are *disrupted* by the inhibition of RNA or protein synthesis.

Among the *correlational* studies, one of the most notable was that of Hydén and Egyházi (1964), which showed that the quantity of RNA, and the ratio of its nucleotide bases, covary in cortical pyramidal neurons with the acquisition of a new information. The authors discuss the hypothesis that the information might be stored in the neurons by means of a molecular mechanism involving the formation of a stable base-permutation of RNA. Proposals such as these led some researchers to speculate that if memories were encoded in specific macromolecules, be they RNA or proteins, it should be possible to isolate and transfer them among individual organisms. Jacobson *et al.* (1965) reported such a transfer of RNA from trained to untrained rats, claiming that the untrained rats subsequently showed the learned behavioural features. Even more extraordinary were the reports of McConnell (1962) that untrained worms, fed with ground-up trained worms, exhibited the trained behaviours. These experiments could not be reliably replicated; Byrne *et al.* (1966) argued convincingly against Jacobson's results, and McConnell later confessed himself a scientific humorist (1971). A side-effect of the memory-transfer experiments was, however, in the longer term, to bury the hypothesis that memories are encoded in either RNA or proteins.

The studies of the *disruption* of learning and memory by inhibition of RNA or protein synthesis were more successful. Initially, the studies focussed on protein synthesis inhibition, and worked with a protein-coding hypothesis (*e.g.*, Flexner *et al.* 1967; Agranoff *et al.* 1965). Flexner *et al.* are especially explicit in elaborating the hypothesis. They link the formation and maintenance of cognitive states closely to the formation and maintenance of new patterns of genetic expression (as in cell differentiation), but they do not attempt to identify the two, mainly because of their finding that protein-synthesis inhibitors disrupt learning and memory. In view of this, they propose that the establishment of a memory depends on the formation of a self-sustaining system involving the genes, RNAs, and their protein products, with the proteins being the key items in 'memory traces'. These studies were not free from alternative interpretations, and in later research the macromolecular theories were all but abandoned (for a review, see Agranoff 1978). However, the Flexner *et al.* hypothesis deserves a special mention as the most sophisticated and best-founded among the 1960's proposals, and can be regarded as a precursor of GTM.

I will close this chapter with a few rather general suggestions why, in their time, the macromolecular hypotheses were not successful. To begin with, the quest for a molecular code of cognition was an implicit revival of 19th-century associationism (as of Cajal); for the macromolecules carried in cellular protoplasm were not assumed to have conceptual and propositional structure, and complex mental states were presupposed to be formed merely by the association of such macromolecules *via* neural connections. The trouble with associationism is that it fails to distinguish semantically simple mental terms from complex terms produced by a generative mechanism, the latter from sentences generated by another mechanism, and the sentences from cognitive and emotive operations on their tokens. Hence it fails, as regards the implementation of psychological states in the brain, to seek molecular substrates that could sustain a complex system of symbols, generative mechanisms and cognitive-emotive operations, resting content with an RNA- or protein-coding hypothesis. From the classical point of view, neither RNA nor proteins could do the job; though they are large sequential molecules, they could not accommodate the requirements of CTM. In contrast, nuclear DNA has the needed complexity, and has the added advantage of mechanisms known to maintain stable patterns of gene expression for periods lasting up to the life-time of an organism.

Further, the biology of genes was far less advanced in the 1960's than it is today. Present studies of the disruption of learning and memory by RNA- and protein-synthesis inhibitors are more precise, and focus on the inhibition of third-messenger RNAs and proteins, the function of which is to determine the patterns of expression of late-effector genes *vis-à-vis* learning stimulations. What seemed to Flexner *et al.* (1967) to require assigning a key role to proteins in memory formation, requires only the role of third messengers between immediate-early and late-effector genes. In general, current evidence indicates that patterns of DNA expression are more flexible with respect to experience, the acquired states more stable, their complexity more appropriate to the complexity of cognition, than could have been supposed in the 1960's.

Finally, the revival of Cajal-style associationism was aborted for the mundane reason that most cognitive biologists worked within connectionism, and it was irresistible for them to give away the search for mental symbols in favour of synaptic-*cum*-behavioural mechanisms which did not demand such posits; the more so considering the chronic dearth of understanding of the classical theory among connectionists and associationists. (For example, see Thompson & Donegan (1986); the authors, following Lashley (1950), mould Descartes as a connectionist, and quote Locke, with Hume, as an associationist.) To the connectionists, even associationism is too symbolic, not enough behaviouristic and hence — in the worst sense of the word — not enough *scientific*.

The classical theory has also had a revival of interest since the 1950's and 60's, but in psychology, linguistics, philosophy, and artificial intelligence; not in the biology of cognition. Yet recent advances in the biology of cognition could well substantiate the claims and posits of the classical theory, provided, as I have argued, one abandons the computer analogy and adopts GTM. The evidence noted and acknowledged, we should nevertheless be aware that GTM is a conjecture, and that a sceptical attitude is appropriate in judging it. However, fairness would be served if *connectionism* and, for that matter, *classical computationalism* were judged in the like spirit. For despite the attention they have received, these doctrines have not been able to overcome the most basic objections: classical computationalism still cannot cope with problems of biological implementation; and it is still to be wondered, to put it mildly, whether synaptic connectionism is a *cognitive* theory rather than merely an account of brain function at the level of neural connections. Against such a background, I suggest that a dose of lateral thinking may be just what is needed; and in that respect I recommend GTM. May it prove useful to philosophers and scientists concerned with the mind-body problem, and as pleasant to contemplate as it has been to me.

Chapter 9

The Classical Theory of Mind III

9.1 Toward an Integrated Account of Mind

The goals of this chapter are, firstly, to develop the formal model of CTM further, especially with respect to the mind's *a priori synthetic* propositions and *a priori* synthetic knowledge; secondly, to integrate the psychic-cell ontology and architecture of Chapter 8 into the broader CTM scheme, as set out in Chapter 7, to obtain a unified account of the mind; and thirdly, to show how the formation of the complex symbolic system of an individual mind is determined by the mind's socio-linguistic and physical environment, and how its external, public symbols derive their significance from its internal symbols. Accordingly, there will be three sections: *(i)* on the symbolic system, representation, and cognitive processes of a single, deep-layer, long-term-store psychic cell; *(ii)* on the mind as a system of psychic cells; *(iii)* on the mind's public affairs, linguistic and other. In Chapter 10, I will use the *a priori* synthetic and analytic methods to resolve the so-called "Russell's paradox", which has been characteristic of the Analytic tradition, and the treatment of which clearly separates CTM from that tradition.

9.2 The Symbolic System of a Deep-layer, Long-term-store Psychic Cell

In Sections 8.4.3.1–8.4.3.2, I described a psychic cell as a neural cell which is specialised for the function of mental representation, and which forms, stores, and processes representations as patterns of expression of its genetic material. There are three layers of psychic cells: input or surface-layer cells, each carrying a single semantically simple empirical term; middle-layer cells, each carrying some part of the empirical basis of simple terms, together with a generative mechanism for the production of semantically complex terms; and deep-layer cells, each carrying the entire basis of simple terms, together with a mechanism for the production of sentences from simple and complex terms, and the basis of simple operations on tokens of sentences, or propositions, together with a mechanism for the production of complex sentential operations. The three-layer, vertical system structures each of the four horizontal faculties: the long-term store, the memory former, the working memory, and the sensorium. In this section, I will describe the symbolic system, representation, and cognitive processes of a single psychic cell; and I will focus on a deep-layer, long-term store psychic cell, which is the most complex of psychic cells, capable of forming very complex and stable cognitive-genetic modules. I will divide the task into three parts, dealing with the *syntax*, the *semantics*, and the *epistemology* of the psychic cell.

9.2.1 Syntax.

The deep-layer, long-term store psychic cell contains the following symbolic system:

(i) A finite basis of *semantically simple empirical ideas*, or (tokens of) mental terms. In the formal model, I assumed that there are twenty types of simple empirical idea, signified by the first twenty letters of the alphabet: $A, B, C, \ldots, T$.

(ii) A generative mechanism comprising operations for the production of infinitely many *semantically complex ideas* from the empirical basis. These operations are idea-laden, in that they require *a priori*, non-empirical ideas in order to function. In the model, I assumed that among these *a priori* ideas are the simple ideas $\wedge$, $\neg$, and $=$, working to generate complex ideas such as $(Fx \wedge Gy)$, $\neg Gz$, *etc.*

(iii) A generative mechanism comprising operations for the production of infinitely many *propositions* from the stock of simple and complex ideas. These operations are also laden with *a priori*, simple and complex ideas. In the model, I assumed that there is one such simple *a priori* idea, Σ; and that the mechanism generates three kinds of proposition: identities such as $a=a$, $a=b$; quantified propositions of the form $(\Sigma\delta)\Gamma\delta$; and compounds of the form $\neg\alpha$, $(\alpha \wedge \beta)$, *etc.*

(iv) A finite number of *basic psychological operations* on propositions. These operations are either cognitive or emotive. The basic cognitive operations are epistemic and evaluative; they are operations by which the mind confirms or disconfirms propositions. In the model, I have so far specified five basic evaluative operations, AO_1–AO_5, which are analytic operations, working in analytic rational processes. In Section 9.2.3, I will introduce synthetic operations, and define synthetic rational and other processes. A notable aspect of the cognitive operations is that they are, like the generative operations for complex ideas and propositions, idea-laden, involving the *a priori* ideas of *epistemic values*, of *modal properties*, and *individual constants*.

(v) A generative mechanism for the production of *complex psychological operations* on propositions. The complex cognitive operations are planning, expecting, *etc.*; among the complex emotive operations are hoping, fearing, and so on. A particular mental state, such as believing that

bats are birds, is an instance of the belief-operation on a proposition that means that bats are birds; and a mental process is a causal sequence of instances of such operations.

Comments. A few remarks intended to prevent misconceptions about the formal model. Firstly, the model is obviously not designed — at this stage of its evolution — to match even remotely the intricacy of the symbolic system of the human mind. It involves only symbols and symbolic operations that enable us to illustrate the syntactic, semantic, and epistemic features of CTM, especially the logic implicit in CTM. However, as a part of an ongoing research programme, the model is to be gradually enriched, to be brought closer to psychological reality. Concerning the empirical basis of ideas, the model should distinguish between sensory and reflective ideas, and to map out at least partially the class of simple empirical ideas; similarly, the model should attempt to map out the structure and function of the *a priori* faculty, including the ideas of space, time, number, causation, the self, representation, values, and other.

Secondly, the model is not designed to follow any of the classical theorists to the letter. CTM is a hybrid account of the mind, dating back at least to Plato and Aristotle, and having received contributions from Augustine, Anselm, Aquinas, Ockham, Descartes, Locke and Kant, to name the most prominent. Despite the differences among them, I think most of the classical theorists would find acceptable the five key assumptions of CTM I have just reviewed, as well as the main aspects of its account of representation and knowledge. One might retort that these could hardly be acceptable to Locke, since he did not distinguish between empirical and *a priori* ideas. Remarkably, however, an attentive reader of the *Essay* will find that Locke did, in fact, draw something like the distinction between empirical and non-empirical or *a priori* ideas, though he did not call it so. (Not only the term "*a priori*", *as applied to ideas*, but also the term "empirical" was introduced by Kant; Locke would not have regarded himself as an empiricist; nor, for that matter, would Descartes have regarded himself as a rationalist; these terms we have inherited from Kant.) He distinguished between, on the one hand, experiential ideas which are occasioned in the mind by — and represent — *specific* sensory or reflective properties (such as **blue**, **sweet**, *etc.*); and, on the other hand, ideas which are *not specific* to any sensory or reflective occasioning experiences, but are universally applicable to diverse sensory experiences, or diverse reflective experiences, or both sensory and reflective experiences (*e.g.*, ideas of existence, unity, duration, extension, consciousness, representation, *etc.*; see (II, V–VIII, *passim*); *cf.* (II, XXI, 73)). The latter ideas, we may say, Locke still viewed as empirical, though not as answering to any specific empirical properties; and his main reason for so regarding them was that he denied that the mind has any ideas prior to, and independently of, some or other experiences. But this need not prevent him from acknowledging that such ideas have a special organisational and epistemic function in the system of the mind, and should be classified apart from *basic* empirical ideas.

I will now comment on a couple of terminological issues the reader may find outstanding. Firstly, the term "proposition" has been used, in modern Analytic Philosophy and formal logic, to stand for sundry things. For example, it has been used in the sense of "sentence", whether public or, less frequently, mental; in the sense of "whatever is expressed by a sentence", or "the meaning of a sentence"; in the sense of "the state of affairs represented by a sentence", with the state of affairs taken to consist of the particulars or sets of particulars referred to by the sentence's constituent terms; more recently, in the sense of "set of possible worlds" (with possible worlds often taken as sets of propositions); and so forth. Locke used the term "proposition" for either *tokens* of sentences, or the sentences themselves (*qua types* of symbol); and either for (tokens of) *mental* sentences, or (tokens of) the *public* sentences expressing the mental sentences. Witness, for example: "The signs we chiefly use, are either *Ideas*, or Words, wherewith we make either mental, or verbal Propositions" (1975: II, XXXII, 19). Again: "The *joining* or *separating* of signs...is what by another name, we call Proposition...: whereof there are two sorts, *viz.* Mental and Verbal; as there are two sorts of Signs commonly made use of, *viz. Ideas* and Words" (IV, V, 2). My own usage is roughly Lockian. I use "proposition" to stand for a *token* of a *mental* sentence, but sometimes — for the sake of convenience — for the sentence itself (*qua* symbol *type*). The usage in the sense of "token of a mental sentence" has several advantages. We can properly regard propositions rather than sentences as bearers of truth-values; for a proposition of the same type can be true on some occasions of tokening, and false on other occasions. We can also say that, though many minds can think the same mental sentence (the same type of proposition), they cannot think the same proposition, since each has its own token, on such and such an occasion, of the symbol type. Finally, we can foresee why, as I said in Section 7.2.4, 'the *a priori* faculty of judgement enables the mind to assess the logical modality of any proposition, and to determine the truth-values of such propositions as may be known with intuitive or demonstrative certainty'. Without going into the deep issues of decidability (which in modern formal logic has not concerned the mental code, but idealised public languages), we can conjecture that the mind is able, in principle, to sort out the logical modality of any of its propositions, by the analytic method described generally in Section 7.5, or by the synthetic method to be described in 9.2.3.1.2 (see also Section 10.3).

Secondly, *truth-values* have been regarded, in modern Analytic Philosophy and formal logic, as *semantic* values; and the semantic properties of sentences have been usually defined by phrases such as "α is *true iff...*". In the semantics of CTM, such formulations must be rejected (*cf.* next section); and, accordingly, truth-values cannot be taken as semantic values.

I have used the term "epistemic value" to refer to truth-values. This reflects the view that, though the truth-value of a proposition is objective, so that the mind has to labour to discover it, it can be discovered only by various means of confirmation or disconfirmation, and these are modes of *epistemic*, not *semantic* evaluation, yielding an epistemic rather than semantic value of the proposition.

9.2.2 Semantics.

The semantics of the symbolic system of the psychic cell rests on three principles:

(i) Meaning is *term-based*, in that the primary bearers of it are mental terms, or ideas; and the meaning of a complex idea or proposition is built from the meanings of its constituent simple ideas.

(ii) The meaning of any symbol, idea or proposition, consists in that the symbol *denotes*, or represents, a certain *universal*: that is, a property, relation or, in the case of propositions, a state of affairs.

(iii) The universals denoted are *nominal*: it is the mind, not the environment, which determines the identity of the property, or relation, or state of affairs represented by a symbol.

In short, the meaning of a symbol, simple or complex, empirical or *a priori*, is the relation of denotation or representation between the symbol and a certain nominal universal. I have expressed this by writing $\mathbb{M}(\#) =_{\mathrm{df}} \mathbb{D}(\#, [\#])$, where # is any symbol, [#] is the universal represented by the symbol, and $\mathbb{M}$ and $\mathbb{D}$ stand for "meaning" and "denoting", respectively.

Comments. The universals denoted by mental symbols are (with some qualifications) *nominal*. The nominality might suggest to some readers that the meanings of symbols, *qua* denotations of nominal universals, are not objective, and that their mind-dependence amounts to variability, arbitrariness, and relativism in logic and knowledge in general. This is not so. The foundation of the mind's cognitive system consists, not in any set of axiomatic principles and rules of inference, as foundationalism is often misunderstood, but rather in that the system comprises a class of basic, semantically simple *empirical ideas*; a class semantically simple *a priori ideas* laden in the *generative mechanisms* for complex ideas and for propositions; and a class of *cognitive-evaluative operations* on propositions, also laden with simple *a priori* ideas. This symbolic foundation is taken to be universal for the human mind, and to constitute an *objective ground* not only for the functioning of the mind, but also for the range of its potential knowledge, including logical knowledge. The objectivity of logic resides in the mind's universal symbolic foundation; and the fact that mental symbols represent *nominal*, not real or *noumenal* properties, does not in the least detract from the objectivity of the foundation.

Some readers may wonder how the mind fixes, nominally, the identity of a property denoted by a mental symbol, and what account can be given of the relation of denotation. These issues are well beyond the scope of this book; the best I can do is to give a hint on each, gathering my remarks so as to cover four kinds of symbol: *(a)* simple empirical ideas; *(b)* simple *a priori* ideas; *(c)* complex empirical ideas (which are mixtures of empirical and *a priori* ideas); and *(d)* propositions.

(a) Simple empirical ideas. Each simple empirical idea has, apart from its syntactic and semantic aspects, a certain discrete *form of consciousness*; and this fixes the identity of the nominal property the idea denotes. For example, the idea **blue** has certain syntactic features; it has a semantic property, or meaning, consisting in its denoting the nominal property [blue]; and the identity of the nominal property [blue] is fixed by the *empirical form of consciousness* of the idea, ⟦ blue ⟧: that is, by the sensation of blue which is associated, essentially and inseparably, with each tokening of the idea **blue**.

The account of the relation of denotation between the idea **blue** and the nominal essence [blue] is *nomic*: **blue** denotes [blue] since instances of [blue] in the mind's environment are nomologically sufficient, under normal psychophysical conditions, to occasion tokenings of **blue**; and such an account may be given whether or not the *real* or *noumenal* nature of any instance of the nominal property [blue] is known, or even knowable.

(b) Simple a priori *ideas.* Similarly, the identity of the property denoted by, say, the simple *a priori* idea ∧, namely, [∧], is fixed nominally by the *a priori form of consciousness* ⟦ ∧ ⟧, which is a non-empirical and — in some sense — pure form of consciousness associated with each tokening of ∧. The account of the relation of denotation between ∧ and [∧] is, again, nomic: ∧ denotes [∧] since instances of [∧] *in a mental system* are nomologically sufficient, under normal psychophysical circumstances, to produce tokenings of ∧. However, since the identity of [∧], being nominal, is fixed solely by the *a priori* form of consciousness ⟦ ∧ ⟧ essentially associated with each tokening of ∧, the mind cannot come to know the *real* or *noumenal* nature of the instances of [∧] merely by thinking or tokening ∧; in fact, it would be a difficult problem in natural philosophy to sort out what, in a physically implemented mental system, the *real* essence of the instances of [∧] consists in.

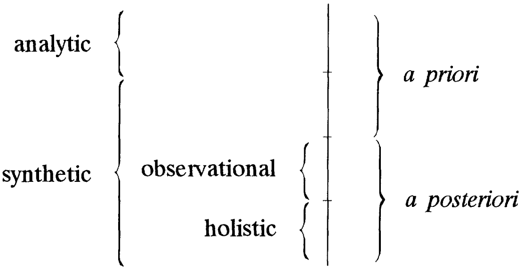
(c) Complex empirical ideas. These are constructions of empirical and *a priori* ideas. The nominality or mind-dependency of the properties denoted by them is therefore ensured by definition; for it is the mind itself which puts such ideas, and hence also the properties they denote, together. The account of the relation of denotation for complex ideas is not nomic (for this applies only to simple ideas); rather, it relies on *semantic construction*, or *definition*. It says that a complex idea, say, $Fx \wedge Gx$, denotes the nominal property $[Fx \wedge Gx]$, since the simple constituent ideas of $Fx \wedge Gx$ — *viz.*, Fx, Gx, and ∧ — denote the simple constituent properties of $[Fx \wedge Gx]$; *viz.*, $[Fx]$, $[Gx]$, and [∧]. In other words, the relation of denotation between a complex idea and the corresponding complex nominal property is reduced

to the relations of denotation between the simple constituents of the idea and the simple constituents of the property. The latter relations are accounted for in *nomic* terms (as in *(a)* and *(b)*).

(d) Propositions. Likewise, the nominality of a *state of affairs*, *qua* universal represented by a proposition, is ensured simply by semantic construction, or definition; for it is the mind itself which organises simple constituent ideas into the complex symbol which is the proposition, and hence also it is the mind itself which organises the properties denoted by the simple ideas into the complex universal which is the state of affairs. Accordingly, the account of the relation of representation between a proposition and the corresponding state of affairs is *definitional*: a proposition represents such and such a state of affairs since the simple constituent ideas of the proposition represent such and such properties; again, the latter relations of representation are accounted for in *nomic* terms. A state of affairs represented by a proposition is therefore a mind-dependent construct of the nominal properties represented by the simple constituent ideas of the proposition (it is certainly nothing like a set of 'possible worlds'.) For example, $\neg a=b$ represents the state of affairs $[\neg a=b]$; and the relation of representation reduces to the relations of denotation between a and $[a]$, b and $[b]$, $=$ and $[=]$, and $\neg$ and $[\neg]$. Further, just as $\neg a=b$ is a mind-dependent construct of a, b, $=$, and $\neg$, so $[\neg a=b]$ is a mind-dependent construct of $[a]$, $[b]$, $[=]$, and $[\neg]$, the identity of which is determined nominally, by the proposition $\neg a=b$, or in general by the mind itself.

9.2.3 Epistemology: cognitive processes in the psychic cell.

We shall next consider the cognitive processes of confirmation or disconfirmation, including implication and inference, in the psychic cell; associative mental processes will be dealt with in Section 9.3. In the tradition of CTM, we shall distinguish broadly two kinds of knowledge acquired by such processes: *a priori* and *a posteriori* knowledge. *A priori* knowledge will be regarded not merely as 'knowledge independent of experience' (as Analytic Philosophers have typically over-simplified and misunderstood it), but rather as knowledge acquired by processes relying solely on *a priori* ideas, with empirical or *a posteriori* ideas being either absent from the proposition known, or merely incidental. Analogously, *a posteriori* knowledge will be regarded not merely as 'knowledge dependent on experience', but rather as knowledge acquired by processes relying *not solely* on *a priori* ideas, but involving crucially *a posteriori* or empirical ideas. These notions will be explained shortly. Further, we shall allow that *a priori* processes can be either *analytic* or *synthetic*, though *a posteriori* processes can be only synthetic; and that synthetic processes which are *a posteriori* can be either *observational* or *holistic-probabilistic*. The following diagram will serve to sketch the picture:



The diagram — rather, the thought it conveys — has a very long history in CTM. Allowing for differences in terminology, it (or a version of it) appears for the first time in Plato (509–511) as 'the simile of the divided line'; and it reappears under various guises with most or all CTM theorists. Currently, the best known version of it is that of Kant; but, as we shall see later, Kant's version is rather idiosyncratic, and should not be accepted as the definitive authority. We shall use the diagram to structure the remainder of this section. In the following sub-section, we shall treat of *a priori* knowledge, and in particular of *a priori* analysis and *a priori* synthesis; then we shall turn to *a posteriori* knowledge, and specifically to observational synthesis and holistic synthesis.

9.2.3.1 *A priori* knowledge.

The standard notion of an Analytic Philosopher that 'the *a priori*' is 'knowledge independent of experience' is vague and misleading, since it obscures or ignores the underlying CTM account of mind: that the mind is an ideational system, in which simple empirical ideas are organised into complex ideas and propositions by operations involving simple and complex *a priori* ideas; and that some propositions their constituent *a priori* ideas constrain to be true, or false, solely by the semantic identity of the *a priori* ideas, and regardless of any other (such as experiential) evidence. In the proofs of these propositions, *a posteriori* or empirical ideas are either absent or incidental (we shall come to examples later). Such propositions, when confirmed either by analysis or synthesis from their constituent *a priori* ideas alone, become *a priori* knowledge. Let us now look at the methods of *a priori* analysis and synthesis in detail.

9.2.3.1.1 Analysis.

Analysis is the process of breaking a proposition down to its simple constituent ideas. This in itself could not reveal either the proposition's modal status, or its epistemic value; the way to find out its logical modality is to analyse it *under an assumption of epistemic value*, seeing whether the clear and distinct semantic identity of its simple constituent ideas accords with the assumption of value: if it does not, then the proposition *cannot* have that value; if it does, then it *may* have it; if it may have either value, then it is contingent. In Section 7.5.1, I set out the analytic method in the formal model of CTM, defining five cognitive-evaluative operations for the *a priori* analytic method of proof:

AO_1 *Assume* the proposition under evaluation has a certain *epistemic value*, either *true* or *false*.

AO_2 *Infer the epistemic value* of each *atomic constituent* of the proposition under evaluation, by the semantical clauses $\Re_{\#}$.

AO_3 *Discern the meaning* of each atomic proposition in terms of its simple constituent ideas, and determine whether each idea has a clear and distinct semantic identity, given the assignment of epistemic value to the atomic proposition.

AO_4 *Judge the epistemic value* of the proposition under evaluation.

AO_5 *Judge the logical modality* of the proposition under evaluation.

These operations, though specified for the simple formal model, can be generalised to apply to any model of the mental code, however complex; for the basic procedure of assuming an epistemic value of the proposition under evaluation, breaking it down to its atomic constituent propositions and assigning a value to each according to the initial assumption, discerning the meaning of the atomic propositions in terms of their semantically simple constituent ideas, judging the epistemic value of the proposition, and judging its logical modality, would be the same for any CTM system. In Section 7.5.2, I gave a couple of examples of *a priori* analytic proof, using the trifling propositions $a=a$ and $(\mathrm{A}x)(Fx \supset Fx)$. We should bear in mind, since we are describing the cognitive processes in a single, deep-layer, long-term store psychic cell, that such simple analytic processes are assumed to occur within the single cell, implemented as sequences of genetic operations on genetic symbols; in Section 9.3, we shall discuss more complex cognitive processes, implemented in systems of psychic cells.

Analytic processes also underlie the mind's capacity to imply the consequent from the antecedent in a true conditional proposition, and to infer the conclusion from the premisses in a valid argument. We shall do well to review these processes in this context.

9.2.3.1.1.1 Implication.

It will be sufficient to look at strong necessary implication. The idea of such implication, $\rightarrow$, denotes an epistemic property of a proposition of the form $(\alpha \rightarrow \beta)$, such that $(\alpha \rightarrow \beta)$ is true *iff* $(\Box\neg(\alpha \wedge \neg\beta) \wedge \nabla\alpha \wedge \nabla\beta)$ is true; in other words, *iff* $(\alpha \wedge \neg\beta)$ is evaluable as false *a priori*, α is evaluable neither as false nor as true *a priori*, and β is evaluable neither as false nor as true *a priori*, by the analytic evaluative operations AO_1–AO_4. The gist of the modality $[\rightarrow]$ is that $(\alpha \rightarrow \beta)$ is true *iff* $(\alpha \wedge \neg\beta)$ is contradictory, in the *ex terminis* sense, but its contradictoriness is due to the *conjunction* of α and $\neg\beta$, not to either of the conjuncts taken severally. The conjuncts as such must be contingent; and the reason for this requirement is that the mind cannot assume a necessarily false antecedent as true without violating the clear and distinct semantic identity of the simple constituent ideas of the antecedent; nor can it imply a necessarily false consequent as true; and since $\Box((\alpha \rightarrow \beta) \equiv (\neg\beta \rightarrow \neg\alpha))$ is true in the mental code, the mind can neither assume a necessarily true antecedent, nor imply a necessarily true consequent, without violating the clear and distinct semantic identity of the simple constituent ideas of the antecedent or consequent. We noted in Section 7.6.2.1 that a number of people in the Analytic tradition have had an intuitive grasp of this requirement; but only CTM can provide an explicit rationale for it.

9.2.3.1.1.2 Inference.

Inferential or deductive mental processes are *proposition-based* rather than term-based, in that they proceed by inferential or deductive rules defined on propositions, rather than *ex terminis*. But the rules of inference are in turn evaluable by *ex terminis* analysis alone, so that inferential processes are reducible to *ex terminis a priori* processes. Each rule of inference is a true, strong, necessary conditional; and a deductive inference is valid just in case there is a way of deducing its conclusion from the premisses by a finite number of applications of the rules. It follows that there cannot be a valid argument with a necessarily true or false conclusion, or a necessarily true or false conjunction of premisses; for no true strong necessary conditional could serve as a deductive rule for it. However, the premisses of a valid argument may be a mix of necessary and contingent propositions; for example, $(\xi \rightarrow \varsigma)$, $\neg\varsigma \vdash \neg\xi$ is a valid argument, with $(\xi \rightarrow \varsigma)$ necessary, and ξ and ς contingent. Contrary to initial appearances, the requirement that the conclusion and the conjunction of premisses in a valid argument be contingent does not impair the demonstrative sciences; for the CTM position is that any necessarily true or false proposition is provable as such by *ex terminis* analysis or synthesis alone, independently of any other evidence, including any evidence one might allege to draw from other propositions, taken as premisses, by rules of inference. To make it very clear, necessary propositions are necessary because they are provable *ex terminis* and *a priori*, from their own constituent *a priori* ideas; not because they are deducible from other propositions regarded as axioms. The purpose of deduction or inference is not to prove necessary propositions, but to proceed — by logical means alone — from premisses which, in their conjunction, are contingent, to contingent conclusions *not yet given or known*: a Sherlock Holmes' rather than a mathematician's enquiry.

9.2.3.1.2 Synthesis.

Kant's version of CTM was perhaps the only one which regarded the mind's representations as wholly unveridical with respect to the noumenal world, and which identified the physical world with the mind's nominal world. Another peculiarity of Kant's version, unquestioningly carried over into modern Analytic Philosophy, was that analysis and synthesis were regarded as *mutually exclusive*, so that an analytically true proposition could not be synthetically true, and conversely. This was in contrast with CTM tradition, as we shall see later. We may note, to begin with, that Kant's proposition "7+5 = 12" is indeed provable by *a priori synthesis*. Supposing its semantically simple constituent *a priori* ideas be that of unity, of addition, and of identity, it is clear that one can synthesise the idea of 7, the idea of 5, the idea of 7+5, the idea of 12, and that one can ascertain the identity between 7+5 and 12. However, it is no less clear that "7+5 = 12" is provable by *a priori analysis*; for assuming the proposition false, we can analyse it down to its simple constituent *a priori* ideas of unity, addition, and identity, and find that the ideas do not have, *under the assumption*, clear and distinct semantic identity, so that the proposition must be true. The reason why Kant believed the proposition *a priori* synthetic, but not analytic, was that he thought of analysis as relying on the *conceptual containment* of predicate in subject in an analytic proposition. But there is no justification for such a restriction; it shows only that Kant did not properly understand the notion of analysis, not that analysis and synthesis must be taken as mutually exclusive. (Besides, strictly speaking, the idea of 7+5 is contained in the idea of 12, and conversely, the idea of 12 is contained in the idea of 7+5; for, in the final analysis, both are complex ideas identical to the complex idea of 1+1+1+ ... +1, twelve times.) We shall henceforth restore the correct traditional understanding of analysis and synthesis, that if a proposition is necessarily true, it is provable as such either by *a priori* analysis or by *a priori* synthesis; and that analysis and synthesis, far from being mutually exclusive, are in fact complementary one to the other: synthesis is a process opposite in direction to analysis, in that whilst analysis proceeds from the proposition under evaluation down to its semantically simple constituent ideas, synthesis works from the simple constituent ideas up to the proposition.

My plan for this section is, firstly, to explain the notion of *a priori* synthesis in more detail, but informally; secondly, to specify four synthetic operations, SO_1–SO_4, in the formal model of CTM, matching roughly the analytic operations AO_2–AO_5; thirdly, to demonstrate by *a priori* synthesis the same trifling propositions, $a=a$ and $(Ax)(Fx \supset Fx)$, which we have already proved by *a priori* analysis in Section 7.5.2; and lastly, to point out some historical precedents of the notion of synthesis and analysis here proposed.

Ideational synthesis, in the most general sense, is the process of putting a proposition together from its simple constituent ideas according to the rules of syntax; thus any proposition whatever may be synthesised (as well as analysed) by the syntactic rules. We are, however, interested in *proving* a necessary proposition by ideational synthesis alone; and this can be achieved provided the synthesis is guided not merely by syntax, but rather by the semantic identity of the proposition's simple constituent *a priori* ideas; in short, by *a priori* synthesis. In this method of proof, we can distinguish roughly four phases: firstly, tokening the simple constituent ideas of the proposition under evaluation, and *discerning* their *clear and distinct semantic identity*; secondly, *inferring the epistemic values* of the atomic propositions, and the constructs of the atomic propositions, which make up the proposition under evaluation; thirdly, *judging the epistemic value* of the proposition (or some of its intermediate propositional stages) *on the grounds of its* a priori *ideas alone*, and regardless of incidental empirical ideas, where there are any; and finally, *judging the logical modality* of the proposition under evaluation. The operations which constitute these synthetic phases are analogous to the analytic operations, only running in reverse. Notably, the analytic operation of *assuming an epistemic value* of the proposition under evaluation (AO$_1$) has no counterpart in the synthetic method; and this because analysis, unlike synthesis, could not work unless under an assumption of value.

I will next spell out the synthetic operations in the formal model of CTM.

SO$_1$ *Discern the meaning* of each constituent simple idea of the proposition under evaluation.

Comments. I will use the phrase '**DiscernM**(*#*) $=_{\mathbf{df}}$ □(*#*, [*#*])' for the operation SO$_1$, where # is any semantically simple idea; and I will form a list of such phrases for each simple constituent idea, empirical or *a priori*, of the proposition under evaluation.

SO$_2$ *Infer the epistemic values* of the *atomic constituents*, and constructs of atomic constituents, of the proposition under evaluation, by the semantical clauses $\Re_{\#}$.

Comments. SO$_2$ will be written as '**InferV**(α) = τ/φ ($*$, $\Re_{\#}$)', where α is any atomic proposition, or a truth-functional construct thereof, of the proposition under evaluation, $*$ stands for the number(s) of the line(s) from which the inference is drawn, and $\Re_{\#}$ indicates the semantical clause(s) by which it is drawn. In the case of basic empirical ideas tokened as predicates in the operation SO$_1$ — such as *Fx*, *Gy*, *etc.* — SO$_2$ will firstly draw a *particular inference*, as in **InferV**((Σx)*Fx*) = τ, and then instantiate to an individual constant; for example, **InferV**(*Fu*) = τ. As such, the operation SO$_2$ is laden with the *a priori* ideas of truth-values (*i.e.*, the simple idea τ and the derived idea φ), and with individual constants. These ideas are the same as those laden in AO$_1$–AO$_5$, and are given the same semantical clauses (see 7.5.1). Also, SO$_2$ will draw inferences of, and assign truth-values to, truth-functional compounds of the atomic propositions, by the semantical clauses for the *a priori* ideas of truth-functions.

SO_3 *Judge the epistemic value* of the proposition under evaluation (or some of its intermediate propositional stage) *on the grounds of its constituent* a priori *ideas alone*, and regardless of incidental *a posteriori* ideas, where there are any.

Comments. I will write SO_3 as '**JudgeV**(α) = T/Φ/N', where α is the proposition under evaluation (or one of its intermediate propositional stages), T is an *a priori* idea standing for the epistemic value *true* when that value is judged on the grounds of the constituent *a priori* ideas alone, Φ is an *a priori* idea standing for the value *false* when that value is judged on the grounds of the *a priori* ideas alone, and N denotes not a truth-value but indicates that SO_3 is unable to determine the truth-value from *a priori* ideas alone. The ideas T and Φ are not new ideas introduced into the mental code; they are simply the ideas of truth and falsehood respectively, *when truth or falsehood is judged solely from* a priori *ideas*. For example, the compound ($Fu \supset Fu$) is judged to have the value [T], since its truth depends solely on the meaning of the *a priori* idea $\supset$, and not at all on the incidental empirical idea Fx. Nor should T and Φ be regarded as ideas of *necessary* truth and falsehood, respectively. Again, they are ideas of truth-values, tokened in SO_3 when the values are judged from *a priori* ideas alone; one might picture SO_3 as running through the two combinations of values of Fu in ($Fu \supset Fu$), judging that ($Fu \supset Fu$) is [T] because it is [τ] for each of the combinations, so that the idea Fx is incidental and only the meaning of $\supset$ matters.

Lastly, when SO_3 assigns the value [T] to a proposition α, it will also assign [T] to the *universal closure* of α; when it assigns [Φ] to α, it will assign [Φ] to the *particular closure* of α; otherwise it will assign [N] to *either closure* of α. For instance, if **JudgeV**($Fu \supset Fu$) = T, then also **JudgeV**($(\mathrm{A}x)(Fx \supset Fx)$) = T; if **JudgeV**($Fu \wedge \neg Fu$) = Φ, then **JudgeV**($(\Sigma x)(Fx \wedge \neg Fx)$) = Φ; and if **JudgeV**(Fu) = N, then both **JudgeV**($(\Sigma x)Fx$) = N, and **JudgeV**($(\mathrm{A}x)Fx$) = N.

SO_4 *Judge the logical modality* of the proposition under evaluation.

Comments. I will use the phrase '**JudgeMOD**(...α) = τ/φ' for the operation SO_4, where α is the proposition under evaluation, and '...' stands for the *a priori* ideas of logical modalities laden in the operation: namely, $\Box$ for necessity, $\Diamond$ for possibility, and ∇ for contingency.

The operation is to work thus:

if **JudgeV**(α) = T, then **JudgeMOD**($\Box\alpha$) = τ;
if **JudgeV**(α) = Φ, then **JudgeMOD**($\Diamond\alpha$) = φ;
if **JudgeV**(α) = N, then **JudgeMOD**($\nabla\alpha$) = τ.

The *a priori* ideas $\Box$, $\Diamond$, and ∇ are the same as those laden in AO_5; in other words, logical modality is the same whether the mind runs a proof by analysis or by synthesis. However, it will be worthwhile to reiterate the semantical clauses for these ideas, couching them this time in the synthetic method, and making it clear that *a priori* synthesis, like *a priori* analysis, is an *ex terminis* method of evaluation, drawing evidence solely from the clear and distinct semantic identity of the simple *a priori* ideas comprised in the proposition under evaluation:

$\mathfrak{R}_{\Box}$ $\mathbb{M}(\Box) =_{\mathrm{df}}$ Д$(\Box, [\Box])$, where $[\Box]$ is an epistemic property of a proposition of the form $\Box\alpha$, such that $\Box\alpha$ is true *iff* α is evaluable as true *ex terminis* and *a priori*, by the evaluative operations SO_1–SO_3 (with α any proposition).

$\mathfrak{R}_{\Diamond}$ $\mathbb{M}(\Diamond) =_{\mathrm{df}}$ Д$(\Diamond, [\Diamond])$, where $[\Diamond]$ is an epistemic property of a proposition of the form $\Diamond\alpha$, such that $\Diamond\alpha$ is true *iff* $\neg\Box\neg\alpha$ is true, *iff* α is not evaluable as false *ex terminis* and *a priori*, by the evaluative operations SO_1–SO_3 (with α any proposition).

$\mathfrak{R}_{\nabla}$ $\mathbb{M}(\nabla) =_{\mathrm{df}}$ Д$(\nabla, [\nabla])$, where $[\nabla]$ is an epistemic property of a proposition of the form $\nabla\alpha$, such that $\nabla\alpha$ is true *iff* $(\neg\Box\neg\alpha \wedge \neg\Box\alpha)$ is true, *iff* α is evaluable neither as true nor as false *ex terminis* and *a priori*, by the evaluative operations SO_1–SO_3 (with α any proposition).

Let us now prove by *a priori* synthesis the same trifling propositions $a{=}a$ and (A*x*)(*Fx* ⊃ *Fx*), which we have already proved by *a priori* analysis.

Firstly, (A*x*)(*Fx* ⊃ *Fx*):

1. **DiscernM**(*Fx*) $=_{\mathrm{df}}$ Д(*Fx*, [*Fx*]), where *Fx* is a basic predicate token, and [*Fx*] is the empirical mode represented ($\mathfrak{Z}$).
2. **DiscernM**($\neg$) $=_{\mathrm{df}}$ Д$(\neg, [\neg])$, where $[\neg]$ is an epistemic property of a proposition of the form $\neg\alpha$, such that $\neg\alpha$ is true *iff* α is not true, *iff* it is not the case that α (with α any proposition) ($\mathfrak{R}_{\neg}$).
3. **DiscernM**($\wedge$) $=_{\mathrm{df}}$ Д$(\wedge, [\wedge])$, where $[\wedge]$ is an epistemic property of a proposition of the form $(\alpha \wedge \beta)$, such that $(\alpha \wedge \beta)$ is true *iff* both α is true and β is true, *iff* it is the case that α and it is the case that β (with α, β any propositions) ($\mathfrak{R}_{\wedge}$).
4. **DiscernM**(Σ) $=_{\mathrm{df}}$ Д$(\Sigma, [\Sigma])$, where $[\Sigma]$ is an epistemic property of a proposition of the form $(\Sigma\delta)\Gamma\delta$, such that $(\Sigma\delta)\Gamma\delta$ is true *iff* some item, in the domain of evaluation of $(\Sigma\delta)\Gamma\delta$, partakes of $[\Gamma]$ (with δ a variable, and $\Gamma\delta$ any predicate-token) ($\mathfrak{R}_{\Sigma}$).
5. **DiscernM**(τ) $=_{\mathrm{df}}$ Д$(\tau, [\tau])$, where $[\tau]$ is the epistemic value *true* of a proposition α, such that α is true *iff* it is the case that α ($\mathfrak{R}_{\tau}$).
6. **DiscernM**(*u*) $=_{\mathrm{df}}$ Д(*u*, [*u*]), where *u* is an individual constant, and [*u*] is the item represented ($\mathfrak{R}_{\mathrm{IC}}$).
7. **InferV**((Σ*x*)*Fx*) = τ (1, 4, 5).
8. **InferV**(*Fu*) = τ (6, 7).
9. **InferV**($\neg$(*Fu* $\wedge$ $\neg$*Fu*)) = τ (2, 3, 8).
10. **InferV**(*Fu* ⊃ *Fu*) = τ (9, $\mathfrak{R}_{\supset}$).
11. **JudgeV**(*Fu* ⊃ *Fu*) = T (10).
12. **JudgeV**((A*x*)(*Fx* ⊃ *Fx*)) = T (11, $\mathfrak{R}_{\mathrm{A}}$).
13. **JudgeMOD**($\Box$(A*x*)(*Fx* ⊃ *Fx*)) = τ (12).

Line 11, **JudgeV**($Fu \supset Fu$) = T, is not merely a rephrasing of Line 10, **InferV**($Fu \supset Fu$) = τ. Rather, the judgement made in 11 is that the truth of ($Fu \supset Fu$) depends on *a priori* ideas alone, and that the truth of Fu, and thus the empirical idea Fx, are only incidental. It might be, for example, that both **InferV**(Fu) = τ and **InferV**(Gu) = τ, so that **InferV**($Fu \supset Gu$) = τ; but this would not justify **JudgeV**($Fu \supset Gu$) = T, since the truth of ($Fu \supset Gu$) depends on the truth of Fu and Gu, and therefore on the *a posteriori* ideas Fx and Gx. Such a situation would justify only **JudgeV**($Fu \supset Gu$) = N; but then, **JudgeV**($(\mathrm{A}x)(Fx \supset Gx)$) = N, and **JudgeMOD** could attribute only contingency to $(\mathrm{A}x)(Fx \supset Gx)$. In general, whenever a synthesis relies crucially on empirical, *a posteriori* ideas, **JudgeV** must assign [N] to the proposition synthesised, and the logical modality of the proposition can be only contingency.

We shall next turn to $a=a$.

1. **DiscernM**(a) $=_{\mathrm{df}} \mathit{\Pi}(a, [a])$, where a is a basic idea, and $[a]$ is the empirical mode represented ($\mathfrak{Z}$).
2. **DiscernM**($=$) $=_{\mathrm{df}} \mathit{\Pi}(=, [=])$, where $[=]$ is a relation between any items ζ and η, such that ζ is identical to η ($\mathfrak{R}_=$).
3. **DiscernM**(τ) $=_{\mathrm{df}} \mathit{\Pi}(\tau, [\tau])$, where $[\tau]$ is the epistemic value *true* of a proposition α, such that α is true *iff* it is the case that α ($\mathfrak{R}_\tau$).
4. **InferV**($a=a$) = τ (1–3).
5. **JudgeV**($a=a$) = T (4).
6. **JudgeMOD**($\Box a=a$) = τ (5).

In a similar way, any necessarily true or false proposition can be proved as such by *a priori* synthesis; equally, it can be proved by *a priori* analysis. The view that analysis and synthesis are not mutually exclusive, and that they do not differ essentially but only in the *direction of proof* (and perhaps in the level of abstraction required, as I will point out shortly), is not new. Descartes (1984: 110–111*ff*) discusses the methods of analysis and synthesis, and mentions the distinction between *a priori* and *a posteriori* confirmation. His use of the terms "*a priori*" and "*a posteriori*" is not quite clear in that context, but his account of analysis and synthesis is clear enough. He says that the two methods are the reverse of one another, with synthesis working from basic notions up to basic propositions, and hence up to the proposition which is to be proved, whilst analysis working from the proposition down to its constituent parts. Likewise, Plato (509–511) describes two methods of abstract proof, as in geometry, the one proceeding from constituent parts up to what is to be shown (synthesis), the other proceeding from what is to be shown down to the simple constituent parts (analysis). Plato and Descartes both emphasise that synthesis is more suitable for teaching what has already been discovered, than it is for discovery itself; whereas analysis is the proper way of discovery but, being more abstract, it is less suitable for teaching. Both also regard synthesis, in contrast to analysis, as lending itself to — so to speak — *empirical exhibition*. For example, in proving a geometrical proposition by synthesis (as Socrates does, or has a servant to do, with a simple version of Pythagoras' theorem in the *Meno*), one can draw pictures of the component parts, using them as aids in the proof; again, in demonstrating that $7+5 = 12$ by synthesis, one can resort to an abacus as an empirical crutch to get the result. Analysis, though, is a purely intellectual way of breaking a proposition down to its simple component parts, and as such does not lend itself to empirical exhibition.

9.2.3.2 *A posteriori* knowledge.

A posteriori knowledge is of a proposition the evaluation of which depends not solely on its constituent *a priori* ideas, but also on something else: either on its constituent *a posteriori* or empirical ideas (in addition to its *a priori* ideas) and nothing further; or on evidence drawn from other propositions already established to some degree of probability, as well as on its constituent *a posteriori* and *a priori* ideas. In the former case, the evaluation and knowledge is (apart from the ever-present *a priori* component) purely *observational*; whilst in the latter case, it is *holistic-probabilistic*. The term "holistic" suggest that it is a matter of pragmatic decision which, and how many, other propositions are brought to bear on the evaluation, and that the other propositions may come from any branch of learning; the term "probabilistic" suggests that such evaluation is always a matter of likelihood, not of conceptual or even observational certainty. The logical status of either observational or holistic knowledge is *contingency*; and the way of confirmation involved is *synthesis* (analysis is not applicable here). The difference between *a priori* synthesis and *a posteriori* observational or holistic synthesis lies in the source of evidence. In the case of *a priori* synthesis, the evidence is drawn solely from the clear and distinct meaning of the simple constituent *a priori* ideas of the proposition under evaluation; in the case of *a posteriori* observational synthesis, the evidence is drawn, in addition, from the simple constituent *a posteriori* ideas; in the case of *a posteriori* holistic synthesis, it is drawn, in addition to the foregoing, from other propositions already more or less established. Purely observational synthesis, like *a priori* synthesis, is *ex terminis*, in that the evidence comes from the constituent simple ideas, *a posteriori* and *a priori*, of the proposition under evaluation. By contrast, holistic synthesis is not *ex terminis*, in that the sources of evidence lie, in part, outside of the proposition's constituents. Most pieces of knowledge people commonly claim are of the holistic sort; and whether this should count as knowledge is open to some doubt, as we shall see later. Purely observational knowledge is commonplace (despite what the modern and post-modern holists would have us believe), insofar as tokenings of *a posteriori* ideas are often *sufficient* evidence for the evaluation of a proposition; we shall accordingly speak of *observational certainty*, as distinct from conceptual, *a priori* certainty on the one hand, and holistic likelihood on the other. We shall deal with knowledge by observational synthesis first, setting out a couple of examples; then we shall briefly look at holistic synthesis.

9.2.3.2.1 Observational, *a posteriori ex terminis* synthesis.

It can be easily shown in our model that some propositions, though contingent, are nevertheless confirmable solely *ex terminis*, by observational synthesis. Supposing the idea *Fx*, say **blue**, is tokened, the proposition $(\Sigma x)Fx$, or **some thing is blue**, can be synthesised with observational certainty as follows:

1. **DiscernM**$(Fx) =_{\mathrm{df}} \mathbb{I}(Fx, [Fx])$, where *Fx* is a basic predicate token, and $[Fx]$ is the empirical mode represented $(\mathfrak{Z})$.
2. **DiscernM**$(\Sigma) =_{\mathrm{df}} \mathbb{I}(\Sigma, [\Sigma])$, where $[\Sigma]$ is an epistemic property of a proposition of the form $(\Sigma\delta)\Gamma\delta$, such that $(\Sigma\delta)\Gamma\delta$ is true *iff* some item, in the domain of evaluation of $(\Sigma\delta)\Gamma\delta$, partakes of $[\Gamma]$ (with δ a variable, and $\Gamma\delta$ any predicate-token) $(\mathfrak{R}_{\Sigma})$.
3. **DiscernM**$(\tau) =_{\mathrm{df}} \mathbb{I}(\tau, [\tau])$, where $[\tau]$ is the epistemic value *true* of a proposition α, such that α is true *iff* it is the case that α $(\mathfrak{R}_{\tau})$.
4. **DiscernM**$(u) =_{\mathrm{df}} \mathbb{I}(u, [u])$, where u is an individual constant, and $[u]$ is the item represented $(\mathfrak{R}_{\mathrm{IC}})$.
5. **InferV**$((\Sigma x)Fx) = \tau$ (1–3).
6. **InferV**$(Fu) = \tau$ (4, 5).
7. **JudgeV**$(Fu) = \mathrm{N}$ (6).
8. **JudgeV**$((\Sigma x)Fx) = \mathrm{N}$ (7).
9. **JudgeMOD**$(\nabla(\Sigma x)Fx) = \tau$ (8).

That **JudgeV** on Line 7 assigns the value [N] to *Fu* indicates that **InferV** on Line 6 assigns $[\tau]$ to *Fu* not by *a priori* synthesis alone, but only contingently on the *a posteriori* idea *Fx*. However, the step **InferV**$((\Sigma x)Fx) = \tau$ has already been taken on Line 5, on the evidence of the tokening of the *a posteriori* idea *Fx* (Line 1); and this is *sufficient* to show that $(\Sigma x)Fx$, though contingent, is true. We know its truth with *observational certainty*; this is not *a priori* certainty, but it is certainty none the less, not a mere holistic likelihood. I will not combat here arguments from dreaming and the like; nor need I worry whether empirical ideas are veridical with respect to the natural world. The proposition, as a *token* of a mental sentence, is established on the occasion of the tokening of the idea *Fx*, because the tokening is sufficient evidence for $(\Sigma x)Fx$, and because the operation **InferV**$((\Sigma x)Fx) = \tau$ is something the mind does in response to the tokening of *Fx* as a matter of course, willy nilly. This allows us of a *logical doubt* as to the existence of the natural world, but not — if I may use such a phrase — of a *nomological doubt*. It is worth bearing in mind that the natural or physical world is not knowable by logical means alone; so we should not be surprised at not having a logical certainty of its existence.

Briefly, it should be clear that more complex *a posteriori* knowledge can be had solely by observational *ex terminis* synthesis. For example, $(\Sigma x)(Fx \wedge Gx)$ may be synthesised thus:

1. **DiscernM**(Fx) $=_{\mathrm{df}}$ Д(Fx, [Fx]), where Fx is a basic predicate token, and [Fx] is the empirical mode represented ($\mathfrak{Z}$).
2. **DiscernM**(Gx) $=_{\mathrm{df}}$ Д(Gx, [Gx]), where Gx is a basic predicate token, and [Gx] is the empirical mode represented ($\mathfrak{Z}$).
3. **DiscernM**($\wedge$) $=_{\mathrm{df}}$ Д($\wedge$, [$\wedge$]), where [$\wedge$] is an epistemic property of a proposition of the form ($\alpha \wedge \beta$), such that ($\alpha \wedge \beta$) is true *iff* both α is true and β is true, *iff* it is the case that α and it is the case that β (with α, β any propositions) ($\mathfrak{R}_{\wedge}$).
4. **DiscernM**(Σ) $=_{\mathrm{df}}$ Д(Σ, [Σ]), where [Σ] is an epistemic property of a proposition of the form $(\Sigma\delta)\Gamma\delta$, such that $(\Sigma\delta)\Gamma\delta$ is true *iff* some item, in the domain of evaluation of $(\Sigma\delta)\Gamma\delta$, partakes of [$\Gamma$] (with δ a variable, and $\Gamma\delta$ any predicate-token) ($\mathfrak{R}_{\Sigma}$).
5. **DiscernM**(τ) $=_{\mathrm{df}}$ Д(τ, [τ]), where [τ] is the epistemic value *true* of a proposition α, such that α is true *iff* it is the case that α ($\mathfrak{R}_{\tau}$).
6. **InferV**($(\Sigma x)(Fx)$ = τ (1, 4, 5).
7. **InferV**($(\Sigma x)(Gx)$ = τ (2, 4, 5).
8. **InferV**($(\Sigma x)(Fx \wedge Gx)$ = τ (3, 6, 7).

And so on, *mutatis mutandis*, for other complex *a posteriori* observational propositions.

9.2.3.2.2 Holistic *a posteriori* synthesis.

As regards describing how human minds form complex representational schemes about the natural environment, and how they interact with the environment and each other as cognitive subjects, *a posteriori* holistic synthesis is the most prominent; as regards cognitive certainty, it is the least prominent; for, on the one hand, most synthetic processes the minds undertake *vis-à-vis* the environment are of the holistic sort; but, on the other hand, these processes can afford them at best likelihood, not certainty. The key aspect of such holistic knowledge, or belief, is that the evidence which guides the synthesis of the proposition believed is drawn from without the proposition itself, and that there can be only pragmatic criteria for choosing where the evidence is to come from, when there is enough of it, and how it is to be weighed against other sources of evidence.

The Classical Theory of Mind has always depreciated holistic knowledge, allowing it the status of a mere opinion or belief; especially so in comparison with *a priori* knowledge. There has been, however, a consensus among CTM theorists, bating differences in terminology, that holistic belief — when properly construed — has an *a priori* foundation; in particular, that natural or physical science has a *metaphysical foundation*. The reason is that in synthesising an *a posteriori* proposition as probably true by holistic evidence, the mind must take into account, in addition to the holistic evidence, also purely observational evidence and, more importantly, the conceptual evidence drawn from the proposition's constituent clear and distinct semantically simple *a priori* ideas; and this latter evidence is not negotiable, if the holistic belief is to be well-construed:

it is not a matter of mere pragmatic choice; in other words, the conceptual schemes of physical science have a metaphysical basis which is objectively given and knowable with certainty. This view raises profound issues as to the veridicality of the mind's *a priori*, metaphysical propositions with respect to the natural world, and with respect to being as such. We shall not go into these issues here; we shall instead return to the nitty gritty of CTM in the context of the psychic-cell organisation of the mind, and the genetic implementation of mental symbols and operations in the brain. We have so far described the *a priori* and *a posteriori* cognitive processes which may occur in a single, deep-layer, long-term store psychic cell. Next we shall turn to larger processes, in systems of psychic cells.

9.3 The Mind as a System of Psychic Cells

The psychic-cell organisation of the mind does not have a precedent in CTM. Locke came closest to it when, in the chapter on *Retention* of the *Essay*, he spoke of ideas as very often "rouzed and tumbled out of their dark Cells, into open Day-light, by some turbulent and tempestuous Passion; our Affections bringing *Ideas* to our Memory, which had otherwise lain quiet and unregarded" (II, X, 7). Locke pictures ideas as though they were monks and nuns dormant in their cells, on occasions roused and tumbled out by some tempestuous passion; not an unfitting way to see the matter. Our point of interest is, however, that Locke thought of ideas as being stored in the mind as implicit types of symbol capable of explicit tokening; that he thought there were many tokens of the same symbolic type, located in cells of a sort; and that the mind was a complex system of such inter-connected cells. (It is worth noting that Locke regarded ideas, and the mind in general, as being indeed *located* and *mobile* in space; see (II, XXIII, 18–21).) About two hundred years later, Ramón y Cajal revived the notion of psychic cells, though within the framework of *associationism*, CTM being by then either misunderstood or forgotten. The study of implementation of mind in brain had much advanced since Locke's time, so that Cajal was able to identify psychic cells with certain kinds of neural cells arranged in several cortical layers. Cajal also proposed that ideas are implemented in the cells by some as yet unknown *molecular substrates*; and, in contrast to connectionists, he held that synaptic links function only to establish associations among the ideas stored as molecules within the cells. Cajal was not in a position to identify the substrates with genetic molecules; but he thought that whatever they might be, the substrates (or the ideas they implement) are implicated in the ontogenic development of the neural cells containing them (*cf.* Section 8.5). In Chapter 8, I proposed that ideas and operations on ideas are implemented in psychic cells as genetic symbols and operations, and that the psychic cells are organised, vertically, in at least three layers, and horizontally in at least four faculties. The surface layer consists of cells carrying a *single semantically simple empirical term*; the middle layer consists of cells carrying some part of the basis of simple empirical terms, together with a generative mechanism for the production of *complex terms*; the deep layer consists of cells carrying the entire basis of simple empirical terms, together with a mechanism for the production *sentences* from terms, and the basis of *simple operations* on tokens of sentences, or propositions, with a mechanism for the production of complex propositional operations. The three layer system functions to generate complex ideas and propositions from the empirical basis in response to experience, by transferring simple empirical ideas from the surface layer down to the middle layer to form complex empirical ideas; and by transferring the complex empirical ideas down to the deep layer to form propositions and structures of propositions (such as arguments, stories, *etc.*). The transfer occurs not by moving the ideas, as symbol tokens, from cell to cell and layer to layer, but by *genetic expression*: an activation of a cell causes an expression of the genetic symbol (*i.e.*, a tokening of the mental symbol) the cell carries; this in turn causes the synthesis and transfer of a signal which the cell passes onto the middle- or deep-layer cells to which it is connected; and this signal causes those cells to token type-identical mental symbols by expressing type-identical genes. Once such complex genetic ideas and propositions are formed in the cells, they can be stored up to the life-time of an organism by means of mechanisms akin to those involved in cell-differentiation. (We may note that neural cells do not divide; many die, but those an organism has at the time of its death are among those it had at its birth; so genetically stored memories need not be disrupted by division.) The horizontal faculties function to link an organism cognitively to its environment: there is the *sensorium*, in which the organism receives external or internal inputs, and forms sensory-level ideas; the *memory former*, which stores and processes new ideas and propositions for several weeks or months, gradually transferring them into permanent storage; the *long-term store* for lasting memory; and the *working memory*, which takes ideas from the sensorium for short-term processing, relays them to the memory former, and orchestrates the re-activation of old ideas, propositions and cognitive processes in the long-term store, in response to information received from the sensorium, the memory former, and the long-term store itself. We shall now consider, given such an organisation of the mind, the main sorts of large-scale cognitive process: namely, the acquisition and retention of ideas and propositions; recollection and remembrance; associative (non-rational) processes; and rational processes.

9.3.1 The acquisition and retention of ideas.

The acquisition and retention of a new idea is the acquisition and retention of a new pattern of genetic expression in a psychic cell. In the case of complex ideas and propositions, learning involves the formative processes of transferring, by genetic expression, their constituent ideas from higher level cells. These formative processes are slow, relying on the chemical messenger pathways that link synapses with underlying patterns of gene expression. In an infant organism, we may take it that the formative processes begin at the *surface-layer* cells of the *sensorium*, and work to form more or less complex sensory level symbols in the middle- and deep-layer cells of the sensorium; only gradually they extend, *via* the working memory, to the memory former and the long-term store. To acquire complex representational patterns across all four of the faculties, and especially in the long-term store, is therefore a slow and laborious process; and once such representational genetic patterns are set in the adult organism, there is little scope for changing the general features of the patterns by, so to speak, re-education. Retention works by mechanisms alike to those involved in the maintenance of altered patterns of genetic expression in cell development and differentiation, and may last for the organism's life-time.

9.3.2 The recollection and remembrance of ideas.

The chemical messenger pathways are too slow to function in relating acquired ideas and propositions to incoming experience, or in mediating behaviour. I suggested in Section 8.4.5 that the acquired genetic-symbolic patterns are re-activated, and thus recalled, in a psychic cell by *voltage-dependent processes*, so that whenever the cell reaches a certain level of depolarisation (perhaps a level similar to that required for cell firing), this triggers the genetic states and processes in the cell nucleus, and so the ideational states and processes implemented by them. The question of which cell is re-activated on which occasion (with respect to which experiences or other occurrent thoughts), and hence which ideational states and processes are recalled and remembered on which occasion, is settled by the patterns of *synaptic connectivity* established between the psychic cells during the formative learning processes. More precisely, supposing a proposition is formed in a deep-layer cell of the long-term store by transferring its component complex ideas from the middle-layer cells, and the complex ideas are formed in the middle-layer cells by transferring their component simple ideas from the surface-layer cells, these formative learning processes establish a pattern of synaptic connectivity between the surface-layer cells and the middle-layer cells, and between the middle-layer cells and the deep-layer cell containing the proposition, such that whenever the state of affairs represented by the proposition is later re-instanced in the environment, it will occasion the activation of the same surface-layer cells — or a similar range thereof — and hence the activation of the same middle-layer cells, until the activation spreads *via* the connections to the deep-layer cell, thereby recalling the proposition. Critically, the patterns of synaptic connections among the psychic cells are established by, and depend on, the connections of ideas stored in the cells, not conversely; *i.e.*, it is not that synaptic connections determine the connections of ideas, but *connections of ideas determine*, with more or less precision, *synaptic connections*.

9.3.3 Associative processes.

A like mechanism, of synaptic connectivity among psychic cells, will account for associative processes; here I regard associative processes as *non-rational* processes, in that they need not be guided by evidence, empirical or *a priori*, but occur because tokenings of certain ideas and propositions are causally connected with tokenings of other ideas and propositions, often without any evidential relations among them. Such co-occurrences of rationally disconnected propositions can be clearly explained by the formation of synaptic connections among psychic cells, which do not match evidential relations among the ideas and propositions stored therein. Associative processes so construed are processes in which *causal relations* among cells are broader, more encompassing, than their *evidential relations*; they are processes which occur because causal relations among cells do not wholly respect the relations of ideas stored in them. These non-rational associative processes are parasitic on rational, evidence-guided processes, as they should be according to CTM; for the patterns of synaptic connectivity among psychic cells are established, with more or less accuracy, by the processes of forming complex ideas and propositions from simple empirical ideas, and these are processes guided by the semantic identity of ideas, not connectivity; so rational ideational processes, involving both empirical and *a priori* ideas, are more fundamental than non-rational associative processes. This view sharply contrasts with either associationism or connectionism, which regard rational processes as (if anything at all) *emerging from* connectivity, and associative or connective processes as fundamental; that is, they regard rationality as parasitic on processes which are non-rational. CTM has always held the opposite: the mind is essentially a rational, idea-processing thing; irrational or non-rational processes are due to its imperfections, not due to its constitution.

9.3.4 Rational processes.

The gist of Kant's objection to Hume's associationism is that ideas are organised in the mind not merely by impassioned associations, but in a lawful manner and rationally, by their semantic identity; in other words, that there are *laws of ideas*. This is really what distinguishes CTM from its opponents, whether ancient, early-modern, or the new-fangled post-moderns. I take rational processes to be those guided by the semantic identity of ideas: in most cases, by both empirical and *a priori* ideas, but in some cases solely by *a priori* ideas; such processes are what the mind is for. I treated of simple rational processes, assumed to occur in a single psychic cell, in Section 9.2, looking into *a priori* and *a posteriori* confirmation, analysis and synthesis, observation and holistic judgement. I wish to emphasise only that rational processes on a larger scale occur not in anything like a central processor of the Classical Computational

Architecture, but rather in systems of inter-connected psychic cells, orchestrated by the working memory. This has the advantage of dispensing with merely sequential processing, allowing instead of parallel and distributed processing, something the connectionists have been wont to claim exclusively for themselves. Thus, whilst simple processes occur in single psychic cells, complex processes may be distributed in many cells over several functional areas of the brain; and several complex processes may be under way simultaneously. Also, complex memories may be distributed over several cells and areas; and vitally important memories may well be *copied*, as one makes backup copies of important files, into several cells, and stored at various sites in the brain. (*Cf.* Lashley's troubles with finding where a rat stores its knowledge about food in a maze, which drove him into his mass-action connectionism, a sort of 'rats holism'.)

9.4 The Mind and its External Affairs

I would like to cover three topics in this section: *(i)* the adaptation of the mind, as a system of psychic cells, to its natural and social environment; *(ii)* the mind's public language; and *(iii)* its levels of intellection and education, as they have been traditionally understood in CTM.

9.4.1 Adaptation in the natural and social environment.

The basis of simple *a posteriori* ideas — together with the simple *a priori* ideas laden in the generative operations for complex ideas and propositions, and the simple *a priori* ideas laden in the cognitive operations on propositions — is common and universal to the human mind; similarly, the vertical and horizontal organisation of the mind as a system of psychic cells is common to us as a species. However, different *individual minds*, having somewhat different formative experiences, and inhabiting different natural and social environments, will form different *complex* ideas, propositions, and complex cognitive operations on propositions; and, accordingly, as systems of psychic cells, they will be differently organised at the level of cellular connections. Likewise, different *communities of minds*, each with its own public language, social history and natural environment, will differ in the complex ideas, propositions, cognitive operations, and cellular organisations of their respective individual members. Such differences are the result of the *adaptation* of an individual mind to its social and physical environment. The mind's ability to adapt is, in effect, its ability to conform its complex ideas, propositions, and propositional operations to the requirements of its environment. I will say that individual minds sharing a common social and physical environment undergo a *regimentation* of their psychic cells. Regimentation is a process whereby the psychic cells of individual minds in a community — in particular, the cells carrying complex ideas expressed by a *shared public word* in the language of the community — tend toward the sameness of their descriptive content; that is, regimentation makes the corresponding psychic cells in individual minds, bound by a shared public expression, carry much the same complex symbols. Clearly (and thankfully), regimentation will never be perfect: even in the most conformist of societies, individual minds will be idiosyncratic in the formation of their complex ideas, propositions, and propositional cognitive (as well as emotive) operations; and we may say that human individuality is protected by such inevitable idiosyncrasies.

In modern Analytic Philosophy, though, variability in ideas — from person to person and time to time — has been held against classical mentalism. It has been taken for granted by all parties that mentalism could not be true unless the complex ideas in the minds of individual speakers, bound by a shared public word, were *strictly synonymous*, giving the same meaning to the word, and unless they comprised *necessary and sufficient conditions* for the membership in the extension of the word. In other words, it has been assumed that mentalism could not be true unless individual minds in a linguistic polity were *totally regimented*, both with respect to one another and with respect to the natural environment; for the requirements of idea-word supervenience and of extension-idea supervenience amount to just such a total regimentation. Classical mentalism is, of course, anything but that, as we shall see next.

9.4.2 Mind and language.

What the Radicals in Analytic Philosophy held against mentalism, and what the Conservatives thought must be put up with, was the charge that since complex ideas vary from individual to individual and, for each individual, from time to time, and since they rarely comprise necessary and sufficient conditions for the membership in the extensions of the words used to express them, mentalism as a term-based account of meaning for public language must be false; for, if it were correct, the sameness of meaning among public utterances would require both the sameness of ideas expressed by the utterances, and the determination of extensions by necessary and sufficient conditions. The strategy of the Conservative Party was to show that these demands, impossible though they seem, can in fact be met. But CTM's position is altogether different; as regards words standing for simple ideas, words standing for complex ideas, and statements standing for propositions, it is as follows.

9.4.2.1 Words standing for simple ideas.

The range of simple *empirical* ideas is the same for all minds who have had like experiences; it can differ only for minds who, for instance, have heard tones or tasted relishes others have not heard or tasted. Further, the range of simple *a priori* ideas is the same for all intact minds; so long as they have had any experiences at all, these are sufficient to occasion the entire range of simple *a priori* ideas. Hence a public word, once associated in the mind of a speaker with a simple idea (and associated correctly according to the linguistic conventions of the speaker's community), has a clear and distinct semantic identity, consisting in that the word denotes the nominal property denoted by the idea. Such words have a high degree of semantic uniformity throughout any linguistic community; for they could differ in meaning among speakers, only if some speakers used them to express ideas other than those agreed upon by the community; but if that happens, there are public means of sorting out the misunderstanding.

It is worth mentioning that public languages do not, in general, contain words for each simple idea the mind can form. This holds for simple *a posteriori* ideas, and perhaps also for simple *a priori* ideas; the simple *a posteriori* ideas are too numerous and finely grained to be matched by public words in any natural language; whereas the simple *a priori* ideas require a great deal of self-knowledge and introspective attention: for example, there may be languages without any words for the simple *a priori* idea $\Box$, though the speakers be nonetheless able to apply the idea, in thought, to propositions such as $a=a$.

There cannot be, as a matter of fact, a one-to-one correspondence between the range of simple ideas and public words; many words apparently signifying simple ideas really stand for *categories* of simple ideas, not for any specific simple idea: *e.g.*, "blue" stands not for any single idea **blue**, but for a whole range of simple ideas denoting several hues of blue; "sweet" stands not for a single idea of sweetness, but for a category of ideas which we cannot differentiate publicly, but can introspectively; and similarly for ideas of sounds, smells, pains and pleasures, emotive states, *etc.*

9.4.2.2 Words standing for complex ideas.

Complex ideas, as has been said, may differ from speaker to speaker and, for each speaker, from time to time. The semantical rule here is that a word, *qua token* of a public symbol, stands for the idea, *qua token* of a mental symbol, which the speaker uses the word to express on that occasion. Thus, meaning is *individualistic* and *nominalistic*: it is individualistic, since it is the speaker first and foremost who means, not a community of speakers or their public language; and it is nominalistic, since it is *tokens* of mental symbols, not types, which are the primary bearers of meaning, with tokens of public symbols meaning only insofar as they stand for tokens of mental symbols. Semantic uniformity across a community of speakers and, for an individual speaker, across time, is achieved by the processes of regimentation which the minds of all speakers are subject to; and although regimentation never reaches its *limit* — *viz.*, the sameness of all complex ideas bound by a common public word, for all speakers, at all times — yet it renders the complex ideas alike enough throughout the community for the purposes of successful public discourse.

As for uniformity with respect to what *extensions* individual speakers *refer to* (*i.e.*, as for extension-meaning supervenience), that is not required for meaning; meaning is *denoting* a nominal universal, not *referring* to an extension or a noumenal universal. Nor should such referential uniformity be required for public discourse; for we are discussing human minds: the mind who would mean by referring, rather than merely denoting, would have to be, with respect to everything it can represent, omniscient. For a divinely perfect mind, denotation and reference would coincide, and the natural world would be how it is meant to be. Such a mind could not only find out how the world is by thinking alone, but also make it as it will by thinking as it will. Meaning and knowing would be inseparable for it, and its nominal world would be the noumenal world. But we are not such a mind.

9.4.2.3 Statements standing for propositions.

A proposition is a token of a mental sentence occurring in an individual mind at a given time. When a statement, or token of a public sentence, is used by the individual mind to express the proposition, its meaning is the meaning of the proposition for that mind on that occasion. If the speaker subsequently alters the symbolic composition of that mental sentence, and then uses a syntactically identical statement to express the new proposition, the statement means what the new proposition means. As in the case of complex ideas, semantic uniformity of public statements of identical syntactic form, made by different persons or the same person at different times, is achieved — more or less, but never completely — by the processes of regimentation. Public communication is not impaired by the fact that meaning is individualistic (variable from person to person) and nominalistic (variable from token to token, or time to time). In fact, the variability is precisely what allows us, privately and publicly, to think on our own, to seek truth where it is not yet known, to pursue justice against injustice, or — as the case may be — to follow blindly the trend of our party, to deceive and lie, to oppress and persecute; in short, it is what allows us to use or abuse our minds and languages for our ends.

9.4.3 Levels of intellection and education.

Think again of the classical diagram of the divided line, as given in Section 9.2.3. The upper two quarters represent *a priori* knowledge, the lower two represent *a posteriori* knowledge; the uppermost quarter represents *analytic* knowledge, the lower three *synthetic* knowledge; the lowest quarter represents *a posteriori holistic* knowledge, the next one up *a posteriori observational* knowledge. Corresponding to these *epistemic* divisions are *ontic* divisions, concerning the realms of being. The upper two quarters represent the *intelligible* realm of *abstract objects*, the lower two the *sensible* realm of *concrete objects*.

Of the lower, sensible realm of concrete objects, the observational quarter is matched by middle-sized concrete objects of common sense, chairs and pears, *etc.*; the holistic quarter is matched by concrete objects which are either too small or too big to be observed, and which we posit as *likenesses* of objects which we can observe, in order to explain and predict what happens at the observable level. Thus, for instance, we posit the existence of the Solar system, to be *like* a system of globes with the Sun at the centre and the planets revolving around it, in order to explain the motions of the wandering stars in the sky, and similar observations; and we posit the existence of the physical atom, to be *like* a miniature Solar system with the nucleus like the Sun and the electrons like the planets revolving around it.

Of the upper, intelligible realm of abstract objects, the analytic quarter is matched by abstract objects which we apprehend solely by intellection, without the aid of any public, observable representations: that is, by analytic thought alone; the synthetic quarter is matched by the same abstract objects, only apprehendable with the aid of public representations such as geometric figures, the beads of an abacus, and so forth. For example, we apprehend by pure intellection the numbers 7, 5, 12, and the relations + and =, and discover by analytic thought alone that $7+5 = 12$; and we may apprehend the same numbers and relations with the aid of observable beads on an abacus, and discover the same truth by synthetic thought, using these beads as an empirical exhibition, if need be.

An infant mind begins to form its complex representations at the observational level; it starts with simple ideas in the sensorium, works its way to some complex representational patterns in the sensorium, and slowly proceeds *via* the working memory and memory former to the long-term store, establishing its early propositional memories. For several years, it will be conversant only about middle-sized concrete objects, matching the observational quarter. Later, it will begin to explain and predict certain regularities in its environment by positing likenesses of familiar objects; let us say, it will be led to posit the existence of Santa Clause in order to explain certain regularities in the behaviour of its parents. Eventually, it will come to learn such abstract truths as that $7+5 = 12$; and here, it will be much easier to start by sliding the beads on an abacus, before moving on to discover the truths solely by rumination. Even then, analytic ruminations will be best practised on simple arithmetical and geometrical propositions, which allow of an easy synthetic check-up: if not explicitly on an abacus or a piece of paper, at least in the mind's imagination. Finally, the mind will have to confront propositions which, though provable both by analysis and synthesis, do not lend themselves to any empirical exhibition in their proofs: for instance, that every event has a cause, or that one ought to treat others as one would be treated by them. There, all manner of difficulties will arise: for one thing, not only the proofs cannot be exhibited in experience, they may seem sometimes to conflict with it; for another, one may have practical reasons against accepting such propositions: one might want or need to treat others in a way one would not be treated by them; else, one's cosmological model of the origin of the universe might conflict with the proposition that every event has a cause. To make matters worse, the mind's public languages will probably not be rich enough to allow of perspicuous expressions of such propositions and their proofs. In such cases, pure reason may very well appear to confound us.

CTM's stance is that such pragmatic considerations should not be brought against the truths of pure reason, and that the pragmatic should be guided by the rational and apodeictic, not conversely; in terms of our diagram, the upper two quarters of the *a priori*, metaphysical sciences should organise and order the lower two quarters of *a posteriori*, physical sciences. Though the mind begin, from a developmental point of view, in the observational quarter, its purpose is to journey upward, through *a priori* synthesis to analysis, till it reaches cognitions of pure *a priori* reason, where no observation will serve as aid to certainty; these will include propositions of logic, higher mathematics, geometry, metaphysics, ethics, theology, and even (*pace* the post-moderns) æsthetics; whence the mind is to reverse its course, coming down to the holistic quarter of natural science and pragmatic belief.

Without question, physical sciences and pragmatic beliefs can be successful, useful, and in large parts true, provided they are organised and orchestrated by the demonstrative, metaphysical sciences. It is fair to say that in contemporary natural sciences and in pragmatic know-how, this classical ordering of knowledge has been inverted: considerations of holistic balance take precedence over demonstrative reason, and scientists are asked to quit certainty in favour of overall probabilistic trends. This is even more so in pragmatic disciplines, social sciences and the humanities, where reason, truth, justice, and beauty are regarded as matters of convention, settled more or less by a balance of political power. Academic institutions, in their business of catering for the needs of what they take to be the real world, lead their researchers, teachers and students alike down to the lowest, holistic quarter, where in the absence of metaphysical constraints, likenesses are made at will to account for other likenesses, images are scrapped and replaced with other images according to the latest fashion or wave of political correctness, and all are made to profess allegiance to this, ironically, *one and only* inverted truth, justice, and beauty. Thought is miserable in the holistic nether-world, but it is so easy: anything goes, if the din of one's party is loud enough.

The classical philosophical path to the upper, intelligible world is much tougher and perilous; one walks in silence and solitude, and any the least error may cause a fall; worse, one carries a heavy baggage with provisions from below; one's eyes have to get used to changes from dark to light, and then again, on the way back, from light to dark; still, some will want to go for the splendour of it. 'Tis an old story; the children are sleepy, time to close the book.

'No, no! We are not sleepy, we want to hear The Tale of Russell's Paradox!'

Ah, The Tale of Russell's Paradox? All right, then, one last chapter, but no more giggling. Once upon a Time — in a far-away Continent, beyond seven Seas and seven Mountains — there lived two Philosophers ... [p.t.o.]

Chapter 10

The Tale of Russell's Paradox

10.1 The Misty Origins of Analytic Philosophy

On 16 June, 1902, Russell wrote to Frege concerning the *begriffsschrift* of the new logic:

> I have known your *Basic Laws of Arithmetic* for a year and a half, but only now have I been able to find the time for the thorough study I intend to devote to your writings. I find myself in full accord with you on all main points, especially in your rejection of any psychological element in logic and in the value you attach to a conceptual notation for the foundations of mathematics and of formal logic, which, incidentally, can hardly be distinguished... On functions in particular (sect. 9 of your *Conceptual Notation*) I have been led independently to the same views even in detail. I have encountered a difficulty only on one point. You assert (p. 17) that a function could also constitute the indefinite element. This is what I used to believe, but this view now seems to me dubious because of the following contradiction: Let *w* be the predicate of being a predicate which cannot be predicated of itself. Can *w* be predicated of itself? From either answer follows its contradictory. We must therefore conclude that *w* is not a predicate. Likewise, there is no class (as a whole) of those classes which, as wholes, are not members of themselves. From this I conclude that under certain circumstances a definable set does not form a whole. (Russell in Frege (1980: 130–131))

Were Frege a classical mentalist, he may have replied, concerning the set-theoretic version of the paradox: "I agree without qualification that there is no class of those classes which are not members of themselves. But *why is it a trouble*?"

I will show in this chapter that, from the perspective of the Classical Theory of Mind, this is indeed no trouble either for semantics or logic, or specifically for set theory. However, as regards Frege's (and Russell's) anti-mentalistic *begriffsschrift*, the objection is destructive; for the only alternatives it leaves open are that either the apparently meaningful subject-term "the set of all and only sets which are not members of themselves" has *no referent*, and the apparently meaningful predicate-term "a set which is not a member of itself" has *no extension* (*i.e.*, no set of all and only sets which are not members of themselves); or else, one has to accept as *necessarily true, as well as false*, the proposition that the set of all and only sets which are not members of themselves is a member of itself *iff* it is not a member of itself. It will be important for us to see just why these options are so destructive for Frege's and Russell's project of constructing a logical *begriffsschrift*, and what effect the dilemma has subsequently had upon the development of Analytic Philosophy. To begin with, here is how Frege himself restates the problem:

> Nobody will wish to assert of the class of men that it is a man. We have here a class that does not belong to itself. I say that something belongs to a class when it falls under the concept whose extension the class is. Let us now fix our eye on the concept: *class that does not belong to itself.* The extension of this concept (if we may speak of its extension) is thus the class of classes that do not belong to themselves. For short we will call it the class K. Let us now ask whether this class K belongs to itself. First, let us suppose it does. If anything belongs to a class, it falls under the concept whose extension the class is. Thus if our class belongs to itself, it is a class that does not belong to itself. Our first supposition thus leads to self-contradiction. Secondly, let us suppose our class K does not belong to itself; then it falls under the concept whose extension it itself is, and thus does belong to itself. Here once more we likewise get a contradiction!
>
> What attitude must we adopt towards this? Must we suppose that the law of excluded middle does not hold good for classes? Or must we suppose there are cases where an unexceptionable concept has no class answering to it as its extension? (Frege 1903: 235)

(Frege's term "begriff", translated as "concept", should be read in the sense of "the property which all and only members of an extension have in common", or "the property which defines the extension"; not, as I use "concept", in the sense of "a token of a mental term", or "idea".) Frege then considers the former alternative, of rejecting the law of excluded middle and thus accepting a *paraconsistent set theory*, but finds it untenable; his main reason is that sets, as proper abstract objects (that is, as abstract particulars), must have a well-defined and unique identity: we cannot have a set which both contains and does not contain an item, whether itself or another, and which is therefore both identical and not identical to itself (much as we cannot have a natural number which is both 1 and 2, or both odd and even); and that is a conclusive reason. He next considers the latter alternative, that there are cases where an unexceptionable concept has no class answering to it as its extension. One might wonder how the expression "unexceptionable concept" is to be understood; for the 'concept' (or property) of being a *set that does not belong to itself* is certainly exceptionable, as we shall see later. What is unexceptionable about it, though, is that the *symbol* or *predicate* "a set that does not belong to itself"

— that is, the symbol expressing the 'concept' of being a set that does not belong to itself — is *meaningful*: for we have just *meant* and *understood* it; the perplexing aspect of the symbol is that, albeit meaningful, it does not have a determinate extension; in other words, it does not conform to the *principle of extension-meaning supervenience*. This Frege thinks is a worry; supposing the former alternative is rejected, he says, "then there is nothing for it but to regard *class names* as *sham proper names*, which would thus not really have any reference" (*ibid.*: 236; emphases mine); that is to say, if there were no determinate extensions for set-predicates, then set-names such as "the set of sets which are not members of themselves" would have to be meaningless shams, which Frege finds as unacceptable as the paraconsistent option.

But now, how should one construe the logical *begriffsschrift* to ensure that it does not turn out as true propositions such as "the set of sets which are not members of themselves is a member of itself *iff* it is not"? This is the dilemma Frege's set theory, and in general his project of a logical *begriffsschrift*, runs into:

> There is thus nothing left but to regard extensions of concepts, or classes, as objects in the full and proper sense of the word [*i.e.*, the paraconsistent option is barred]. At the same time, however, we must admit that the interpretation we have so far put on the words 'extension of a concept' needs to be corrected [*i.e.*, corrected whilst saving the meaningfulness of the predicate expressing the concept]. (*ibid.*: 237)

The reader may have noticed that, given these constraints, Frege's dilemma would be solved if he abandoned the Conservative semantical principle of extension-meaning supervenience; and that would be the correct move in any semantical and logical problem, as I have argued in the foregoing chapters. For it would allow him to say that the *predicate-term* "a set which is not a member of itself" is *meaningful*, yet need not have a determinate *extension*, and that the *subject-term* "the set of sets which are not members of themselves" is meaningful, yet need not have any *referent*; and, in the absence of such a determinate extension and referent, the proposition "the set of sets which are not members of themselves is a member of itself" would turn out false, whilst its negation true, so that the proposition "the set of sets which are not members of themselves is a member of itself *iff* it is not" would be false, as required; no such paradoxical propositions would then arise in the logical *begriffsschrift*. (An analogous account could be given of Cantor's term "the set of all sets", *etc.*).

It is interesting that Frege had in fact eventually moved toward rejecting the principle of extension-meaning supervenience; however, the question he never managed to answer was how to construe the meanings of set-terms, if not by extension-meaning supervenience; for he knew of no other criterion of semantic individuation, in a formal language of set theory, except extension-meaning supervenience; and that is precisely where his anti-psychologism, shared with Russell and other Fathers of Analytic Philosophy, had let him down. Frege's set theory was in effect a *formal application of the principle of extension-meaning supervenience*: an *impossible* ground to build on; and once that principle had been shown untenable by the paradox, the set theory, as far as he was concerned, totally collapsed; Frege was great enough a philosopher to be unwilling to put up with patch-work 'solutions'. In his last extant letter, to Hönigswald, on 26.4–4.5., 1925, Frege declares set theory impossible and repudiates his *begriffsschrift*, offering this explanation of the origin of the paradoxes in it. He firstly reviews the paradoxes, and then continues thus:

> These are the paradoxes of set theory which make set theory impossible. Instead of 'set of *F*s', where '*F*' stands for a concept word, we could say just as well 'extension of *F*' or 'class of *F*s' or 'system of *F*s'. The essence of the procedure which leads us into a thicket of contradictions can be summed up as follows. The objects that fall under *F* are regarded as a whole, as an object, and designated by the name 'set of *F*s' ('extension of *F*', 'class of *F*s', 'system of *F*s', etc.). A concept word '*F*' is thereby transformed into the object name (proper name) 'set of *F*s'. This is inadmissible because of the essential difference between concept and object, which is indeed quite covered up in our word languages. Concept words and proper names are exactly fitted for one another... Because of its need for completion (unsaturatedness, predicative nature), a concept word is unsaturated, i.e., it contains a gap which is intended to receive a proper name. Through such saturation or completion there arises a proposition whose subject is the proper name and whose predicate is the concept word, and which expresses that the object designated by the proper name falls under the concept.
>
> ... In such a proposition, concept word and proper name occupy essentially different places, and it is obvious that a proper name will not fit into the place intended for the concept word. Confusion is bound to arise if a concept word, as a result of its transformation into a proper name, comes to be in a place for which it is unsuited... The expression 'the extension of *F*' seems naturalized by reason of its manifold employment and certified by science, so that one does not think it necessary to examine it more closely; but experience has shown how easily this can get one into a morass. I am among those who have suffered this fate. When I tried to place number theory on scientific foundations, I found such an expression very convenient. While I sometimes had slight doubts during the execution of the work, I paid no attention to them. And so it happened that after the completion of the *Basic Laws of Arithmetic* the whole edifice collapsed around me. Such an event should be a warning not only to oneself but also to others. We must set up a warning sign visible from afar: let no one imagine that he can transform a concept into an object. (Frege 1980: 55)

Somewhat nebulously but nevertheless, Frege here gestures toward the semantical view that a meaningful 'concept word' need not have a determinate extension, so that the 'proper name' "the extension of so-and-so" need not have a referent; and that the paradoxes in his logical *begriffsschrift* arise because of the wrong semantical principle that for each meaning-bearing predicate there must be a determinate extension. In his rejection of this principle, Frege came much closer to solving the paradox than either Russell or any of his successors, let alone the paraconsistent post-modernists of today.

As for Frege's warning sign, 'let no one imagine that he can transform a concept into an object', I agree; I take this to express my polemical thesis that meaning does not determine extension. But then, I would add that we set up the further warning sign: 'let no one imagine that he can dispense with the mind, whether in logic, mathematics, metaphysics, semantics, epistemology, or in human affairs generally'; the intellectual and broader cultural history of this century has been a sufficient demonstration of what happens otherwise.

Frege's own anti-mentalistic position rested on a hopelessly poor notion of classical mentalism:

> Logic ... is in no sense part of psychology. Pythagoras' theorem expresses the same thought for all men, whereas everyone has his own images, feelings and decisions, different from everyone else's. Thoughts are not mental entities, and thinking is not an inner generation of such entities but the grasping of thoughts which are already present objectively.
> (1980: 67)

Such mind-less thoughts inevitably led him into the morass; but that was not the end of it. The philosophical movement which grew out of these beginnings, the inappropriately called "Analytic Philosophy", thereafter continued with mindless thoughts to mind-less behavioural holism, and hence to the contemporary *anti-philosophy philosophy*, when hosts of departmental neo-sophists are busy, in tenure, burying philosophy under the sheer volume of 'the text', or post-modern *papier mâché* (*cf.* Couture & Nielsen 1993).

The principle of extension-meaning supervenience is the *immediate* source of Frege's dilemma, underlying as it does his set theory; and he came to understand this late in his life. Subsequent Analytic Philosophers, Wittgenstein and Quine most notably, also identified that principle as a source of trouble, and moved to reject it; but being anti-mentalists — in this unquestioningly following Frege and Russell — they turned to behavioural holism, where logic and mathematics become no more than a recalcitrant convention, with nothing left of the intended rigour, objectivity and universal validity of Frege's logical *begriffsschrift*. This, in summary, is the tale of Russell's paradox, and of Analytic Philosophy; but the tale does have an happy end, as we shall see next.

10.2 Russell's Sophism in the Classical Theory of Mind

In Chapters 7 and 9, it was not necessary for me to distinguish between *using* and *mentioning* the ideas in my formal model; for I had no occasion to use them to state anything regarding how the world is; I spoke only *about* ideas and propositions, saying, for instance, that any proposition of the form $\Diamond(\Sigma x)\Psi x \rightarrow (\Sigma x)\Diamond\Psi x$ is necessarily false (*cf.* 7.6.2.1), and so forth. In what follows, however, I will not only refer to ideas in my model, but also use them to refer to sets, *qua* abstract particulars; and for this reason, I will resort to the quotation marks to indicate when an idea is mentioned rather than used: thus "$\{\xi\,|\,\xi\notin\xi\}$" will refer to the *idea* of the set of sets which are not members of themselves, whereas $\{\xi\,|\,\xi\notin\xi\}$ will purport to refer to that set; again, "$\{\xi\,|\,\xi\notin\xi\}\in\{\xi\,|\,\xi\notin\xi\}$" will refer to the *proposition* that the set is a member of itself, whereas $\{\xi\,|\,\xi\notin\xi\}\in\{\xi\,|\,\xi\notin\xi\}$ will purport to refer to the truth-condition of that proposition.

With this provision, I will introduce two further semantically simple ideas into my formal model of CTM: the simple *a priori* idea of a *set* or *extension*, "$\{\delta\,|\,\Gamma\delta\}$", and the simple *a priori* idea of set-membership, "$\in$"; in addition to these, I will define the semantic identity of the complex *a priori* idea of the empty extension or null-set, "$\varnothing$", and the identity of the *atomic propositions* in which the simple predicate "$\in$" occurs.

$\mathfrak{R}_{\{\delta|\Gamma\delta\}}$ $\mathbb{M}($"$\{\delta|\Gamma\delta\}$"$) =_{\mathrm{df}}$ Д("$\{\delta|\Gamma\delta\}$", $[\{\delta|\Gamma\delta\}]$), where $[\{\delta|\Gamma\delta\}]$ is the *nominal* property of being the set of all and only items δ — whether abstract or concrete, particular or universal — which instantiate the *real* or *noumenal* property of being Γ (with "δ" a variable, and "$\Gamma\delta$" any predicate-token, simple or complex, in which all occurrences of "δ" are free, and all occurrences of other variables, if there are any, bound).

$\mathfrak{R}_{\in}$ $\mathbb{M}($"$\in$"$) =_{\mathrm{df}}$ Д("$\in$", $[\in]$), where $[\in]$ is a *nominal* relation between any item ϐ — abstract or concrete, particular or universal — and a set $\{\delta|\Gamma\delta\}$, such that ϐ belongs to $\{\delta|\Gamma\delta\}$; that is, such that ϐ instantiates the *real* or *noumenal* property of being Γ (with "δ" a variable, and "$\Gamma\delta$" any predicate-token, simple or complex, in which all occurrences of "δ" are free, and all occurrences of other variables, if there are any, bound).

$\mathfrak{R}_{\varnothing}$ $\mathbb{M}($"$\varnothing$"$) =_{\mathrm{df}}$ Д("$\varnothing$", $[\{\delta|\Box\neg(\Sigma\delta)\Gamma\delta\}]$), where $[\{\delta|\Box\neg(\Sigma\delta)\Gamma\delta\}]$ is the *nominal* property of being the set of all and only items δ, such that δ instantiates the *real* property of being Γ, and "$\Box\neg(\Sigma\delta)\Gamma\delta$" is true (with "$\delta$" a variable, and "$\Gamma\delta$" a *complex* predicate-token in which all occurrences of "δ" are free, and all occurrences of other variables, if there are any, bound).

$\mathfrak{R}_{\mathrm{AP3}}$ $\mathbb{M}($"$\delta\in\lambda$"$) =_{\mathrm{df}}$ Д("$\delta\in\lambda$", $[\delta\in\lambda]$), where $[\delta\in\lambda]$ is a *state of affairs*, composed of $[\delta]$, $[\lambda]$, and $[\in]$, such that $[\delta]$ is a member of $[\lambda]$ (with "δ" and "λ" individual constants).

Comments. The idea "$\varnothing$" is a modal one, relying as it does on the modal idea "$\Box$"; and it is a complex idea, comprising the simple ideas "$\{\delta|\Gamma\delta\}$", "$\Box$", "$\neg$", and "Σ". Notice also that "$\Gamma\delta$" must be a *complex* predicate, since there are no simple predicates "$\Psi\omega$" such that "$\Box\neg(\Sigma\omega)\Psi\omega$" is true; that is, "$\Gamma\delta$" must be a predicate such as "$(Fx \wedge \neg Fx)$", "$y \neq y$", and so forth.

The two simple *a priori* ideas "$\in$" and "$\{\delta|\Gamma\delta\}$", added to the formal model of CTM as developed in Chapters 7 and 9, are all that is necessary for the *foundations of set theory*. Contrast this with the various modern attempts to *axiomatise* set theory, and in general with the contemporary notion that the foundations of a formal logical and epistemic system must be *proposition-based* and *deductive*. To the contrary, CTM holds that the foundations of any logical or epistemic system must be *term-based*, comprising a finite basis of *a posteriori* ideas, a generative mechanism (laden with finitely many *a priori* ideas) for complex ideas, another generative mechanism (also laden with *a priori* ideas) for propositions, and a finite number of basic evaluative operations on propositions, likewise laden with *a priori* ideas.

The simple *a priori* ideas "$\in$" and "$\{\delta|\Gamma\delta\}$" I will regard as laden in the generative mechanism for complex ideas. The idea "$\in$" I will take, formally, as on a par with "$=$"; *i.e.*, as a simple idea laden in the generative operation of *comparison* (*cf.* Section 7.2.2). The idea "$\{\delta|\Gamma\delta\}$" I will take as laden in the generative operation of *composition*, and thus, formally, as on a par with "$\neg$" and "$\wedge$"; that is to say, even though the idea "$\{\delta|\Gamma\delta\}$" is semantically simple (like "$\neg$" and "$\wedge$"), it cannot stand as a term on its own, but must combine with some predicate-token, say "Fx", to form the complex idea "$\{x|Fx\}$" (much as "$\neg$" cannot stand on its own, but combines with "Fx" to form "$\neg Fx$"; *etc.*).

The most important feature of the semantic clause for "$\{\delta|\Gamma\delta\}$" is, firstly, that — like any other idea — "$\{\delta|\Gamma\delta\}$" means insofar as it *denotes* a *nominal* property (*viz.*, the nominal property $[\{\delta|\Gamma\delta\}]$); in other words, the meaning of a symbol always consists in the denoting of a nominal property, whether the symbol be of a set or any other; and, secondly, that the property of being Γ, which fixes the identity of the set $\{\delta|\Gamma\delta\}$ (where there is any such set), is *real* or *noumenal* rather than merely nominal. The second feature says that the identity of the extension of a predicate such as "$\Gamma\delta$" is fixed not by what the predicate *denotes* or *means*, but by what it *refers to*; that is, by the *real* property of being Γ, not by the nominal property $[\Gamma]$; and, in turn, this amounts to a denial of the Conservative principle of extension-meaning supervenience: extensions, or sets, are not in general fixed by meaning alone, and conversely, the meaning of a predicate is independent of the predicate's extension, or even of there being any such extension. In short, the meaning of "$\Gamma\delta$" is the denoting of $[\Gamma]$, regardless of the identity, or even existence, of $\{\delta \mid \Gamma\delta\}$; and the identity of $\{\delta \mid \Gamma\delta\}$ is determined by the *real* property of being Γ, not by $[\Gamma]$.

The reader can perhaps foresee what effect this will have on Russell's sophism; for recall that Frege's dilemma was *either* to accept the existence of the extension $\{\xi \mid \xi \notin \xi\}$ only because the predicate "$\xi \notin \xi$" is meaningful (and hence to accept the *paradoxical* conclusion that $\{\xi \mid \xi \notin \xi\}$ both is and is not identical to itself), *or else* to accept that the clearly meaningful predicate "$\xi \notin \xi$", and the subject-term "$\{\xi \mid \xi \notin \xi\}$", are *meaningless*; which alternatives were forced upon him by the principle of extension-meaning supervenience. But we shall come to apply the set theory to Russell's sophism later; for the moment, it will be useful to sort out, in some detail, how *real* properties determine the identity of sets, and hence also how the real properties *referred to* by predicates determine the identity of the extensions of the predicates.

10.2.1 Levels of being.

I will distinguish four levels of reality, and correspondingly four kinds of *real* or *noumenal* property determining the identity of sets: the *metaphysical*, the *physical*, the *psychological*, and the *social* reality.

10.2.1.1 Metaphysical reality.

Metaphysical reality is that which the mind apprehends solely by its *a priori* faculty, and metaphysically real properties are those it *refers to* by its purely *a priori* ideas, simple or complex. We may use the symbol "$\lceil \ldots \rceil$" for metaphysically real properties; for example, the simple *a priori* idea "$\in$" refers to the metaphysically real relation $\lceil \in \rceil$, the complex *a priori* idea "$\neq$" refers to the metaphysically real relation $\lceil \neq \rceil$, and so forth.

Further, metaphysical kinds such as $\lceil \neq \rceil$ and $\lceil = \rceil$, which are referred to by *a priori* ideas, do not change from time to time, from one individual mind to another, or from one instantiation to another, and *perfectly* determine the identity of their extensions, if they have any extensions; thus, $\lceil \neq \rceil$ perfectly determines the identity of $\{\xi \mid \xi \neq \xi\}$ as being the empty set, $\varnothing$; in contrast, $\lceil = \rceil$ perfectly determines that $\{\xi \mid \xi = \xi\}$ does not exist, since otherwise there would be both $\{\xi \mid \xi = \xi\} = \{\xi \mid \xi = \xi\}$ and $\{\xi \mid \xi = \xi\} \neq \{\xi \mid \xi = \xi\}$.

Lastly, in the case of *a priori* ideas (and only in the case of *a priori* ideas), the *referring* to real properties and the *denoting* of nominal properties *coincide*, so that the real properties and the nominal properties are identical; *e.g.*, the predicate "$=$" refers to the metaphysically real kind $\lceil = \rceil$ and denotes the nominal kind $[=]$, but the reference and denotation coincide, so that $\lceil = \rceil$ and $[=]$ are identical. In effect, this is to say that in the case of *a priori* ideas (and only in that case), meaning does determine extension, or else it determines that there is no extension: for, firstly, the extensions of *a priori* predicates are fixed by the metaphysical properties the predicates *refer to*; and, secondly, the metaphysical properties are identical to the nominal properties the predicates *denote* (or mean). To put it emphatically:

with and only with *a priori* ideas, extensions — where there are any — are determined by, and discoverable from, the clear and distinct semantic identity of the ideas; if there are no extensions, this too is determined by, and discoverable from, the clear and distinct semantic identity of the ideas. Equally emphatically, however, *the principle of extension-meaning supervenience does not hold*, since the meaning of an idea does not always necessitate the existence of an extension, nor does the meaningfulness of an idea depend on the existence of an extension. (Specifically, as we shall see later, the predicate-idea "$\xi \notin \xi$" need not, and does not, have an extension — *viz.*, the set $\{\xi \mid \xi \notin \xi\}$ — in order to be meaningful; nor does the subject-idea "$\{\xi \mid \xi \notin \xi\}$" need a referent — again, the set $\{\xi \mid \xi \notin \xi\}$ — in order to be meaningful.)

It further follows that the *truth-condition* of an *a priori* proposition is determined by, and discoverable from, the clear and distinct semantic identity of its constituent *a priori* ideas, and moreover that it exists; for although some constituent ideas of an *a priori* proposition may not have any extensions (such as the *a priori* predicate "$x=x$"), the proposition itself must have a determinate truth-condition, since even the non-existence of an extension is determined by, and discoverable from, the clear and distinct semantic identity of the ideas, which is sufficient to determine the identity of the proposition's truth-condition. (As regards the truth-condition of Russell's sophism, we shall see how it is determined in Section 10.2.3.)

10.2.1.2 Physical reality.

Physical (or natural) reality the mind apprehends and refers to by some of its complex *a posteriori* ideas (such ideas being invariably constructs of both *a priori* and *a posteriori* ideas, since no simple *a posteriori* idea refers to a physical or natural kind, or any real kind whatever). For instance, the complex *a posteriori* idea **gold** refers to the natural kind <gold>, the idea **bat** refers to the natural kind <bat>, *etc.*

Further, physical or natural kinds such as <gold> and <bat>, which are referred to by certain complex *a posteriori* ideas, do (we may take it) vary to some extent from one particular instantiation to another, as well as from time to time, and therefore determine the identity of their extensions *less perfectly* — so to speak, with a lesser degree of reality — than metaphysical kinds; thus, <gold> determines, more or less perfectly, $\{\xi \mid \text{Gold } \xi\}$; <bat> likewise determines $\{\xi \mid \text{Bat } \xi\}$; and so on. Still, since natural kinds have no existence except as inherent commonalities of concrete particulars (so that if the particulars cease to exist, so do the natural kinds), we may say that for each natural kind there is a determinate non-empty extension: namely, the set of particulars which instantiate it; and hence we may say that the *a posteriori* ideas referring to natural kinds always have determinate non-empty extensions, and propositions about natural kinds always have determinate truth-conditions.

Lastly, in the case of these complex *a posteriori* ideas, the referring to real properties and the denoting of nominal properties (or meaning) cannot coincide, since the real properties and the nominal properties cannot be identical; that is, the principle of extension-meaning supervenience is false: meaning alone is not sufficient, with any such ideas, to determine the identity of their extensions.

10.2.1.3 Psychological reality.

Psychological reality is the *introspective* reality of simple and complex ideas, propositions, and propositional cognitive-*cum*-emotive operations, which the mind apprehends both by its *a priori* and *a posteriori* faculty, insofar as it forms *reflective* ideas referring to other ideas, propositions, and operations on ideas and propositions. Thus, using the symbol "≪...≫" for psychological kinds, the idea of an idea refers to the psychologically real property ≪idea≫; more specifically, the idea of the idea **bat** refers to psychological kind ≪the idea **bat**≫; the idea of the proposition **bats are birds** refers to the kind ≪the proposition **bats are birds**≫; the idea of believing that **bats are birds** refers to the kind ≪believing that **bats are birds**≫; and so forth.

In Chapter 8, I suggested that ideas, propositions, and operations on ideas and propositions are physically implemented in the brain as patterns of expression of the genes in certain neural cells; but I did not commit myself to materialism, taken as the doctrine that mental symbols and operations are to be *identified* with these physical states. This accords with my view that psychological kinds are distinct from physical kinds, though any particular mental symbol or operation has a material implementation (*cf.* Section 8.1).

Like physical kinds, we may taken it that psychological kinds do vary to some extent from one particular instantiation to another and from time to time, therefore determining the identity of their extensions *less perfectly* — with a lesser degree of reality — than metaphysical kinds; thus, ≪the idea **blue**≫ determines, more or less perfectly, {ξ|The idea **blue** ξ}, *etc.* Again, since psychological kinds have no being save as inherent common aspects of concrete particulars (so that if the particulars cease to be, so do the psychological kinds), we may say that for each psychological kind there is a determinate non-empty extension: *viz.*, the set of particulars which instantiate it; and hence also that the ideas referring to psychological kinds have determinate extensions, and propositions about psychological kinds have determinate truth-conditions.

Finally, as in the case of the ideas of physical kinds, with ideas of psychological kinds the referring to real properties and the denoting of nominal properties (or meaning) cannot coincide, since the real and the nominal properties cannot be the same; once more, this is to say that the semantic principle of extension-meaning supervenience is false: meaning alone does not suffice, with any ideas representing mental symbols and operations, to fix the identity of their extensions.

10.2.1.4 Social reality.

Social reality the mind apprehends by its *a posteriori* faculty, and socially real properties it refers to by some of its complex *a posteriori* ideas. I will use the symbol " ⌊...⌋ " for social kinds; for example, the idea **table** refers to the social kind ⌊table⌋, the idea **unicorn** refers to the socially real kind ⌊unicorn⌋, *etc.*

Social kinds are *emergent* universals, with the emergence being construed as follows. Each individual mind in a society forms its own complex *a posteriori* ideas such as **table**, or **unicorn**, denoting (and so meaning) such complex nominal kinds as [table], [unicorn], *etc.*; these complex ideas — and, correspondingly, the nominal kinds denoted — may vary from person to person and, for each person, from time to time; however, under social pressures, the individual minds undergo a psychological-*cum*-linguistic *regimentation* (see Section 9.4), such that the various complex ideas of a table (unicorn, *etc.*) — stored in their psychic cells, and associated with the public word "table" — tend toward sameness as to a limit; this limit of regimentation of the ideas (and hence of the nominal kinds denoted by the ideas) may be taken to determine the identity of such emergent social kinds as ⌊table⌋; and the socially real kind ⌊table⌋ may be said to *emerge* from the various nominal kinds [table], denoted by the sundry complex ideas **table** in the individual minds comprising the society.

Further, socially real kinds such as ⌊table⌋ or ⌊furniture⌋, which are referred to by complex *a posteriori* ideas, do vary a great deal from society to society, from time to time, and within each society (and at any time) from one particular instantiation to another; social kinds therefore determine the identity of their extensions *less perfectly* than either physical or psychological kinds, and still less perfectly than metaphysical kinds. For example, ⌊table⌋ determines, more or less, $\{\xi \mid \text{Table } \xi\}$; ⌊furniture⌋ likewise determines $\{\xi \mid \text{Furniture } \xi\}$; and so on. Also, socially real kinds need not be instantiated by any particular: thus ⌊unicorn⌋ has no instantiations, so that the extension $\{\xi \mid \text{Unicorn } \xi\}$ is the empty set, $\varnothing$. Nevertheless, each idea referring to a socially real kind has a determinate extension, whether empty or not; and hence also each proposition about social kinds has a determinate truth-condition.

Lastly, in the case of the complex *a posteriori* ideas referring to social kinds (as with those referring to physical and psychological kinds), reference and denotation never coincide, since the social kinds referred to and the nominal kinds denoted cannot be identical. This is perhaps obvious in regard of such social kinds as ⌊table⌋ or ⌊furniture⌋; for the nominal kinds [table] or [furniture] are so variable and *idiosyncratic* that none can be identified with the socially real kinds ⌊table⌋ or ⌊furniture⌋. But even in regard of, say, ⌊bachelor⌋ — referred to by complex ideas so well regimented that most minds in a society form much the same description **unmarried marriageable male**, thus denoting much the same nominal kind [unmarried marriageable male] — the social kind ⌊bachelor⌋ cannot be identified with any of the individual nominal kinds [unmarried marriageable male]; for they are a different sort of property: the nominal kinds are fixed *individually*, by the complex descriptions in each individual mind; in contrast, the social kind ⌊bachelor⌋ is fixed *socially*, by the processes of psychological-*cum*-linguistic regimentation of the individual descriptions, and hence of the individual nominal kinds denoted by them. Once again, to say that the socially real and the nominal properties cannot be identical, so that reference and denotation cannot coincide for these ideas, implies that the semantic principle of extension-meaning supervenience is false: sole meaning cannot be enough, with any ideas representing social reality, to determine the identity of their extensions.

These four level of reality — the metaphysical, physical, psychological, and social — and the corresponding four sorts of *real* property, are both sufficient and necessary toward the foundations of set theory, *insofar* as the set theory requires an account of the properties determining the identity of extensions and truth-conditions. However, inasmuch as concerns the Russellian claim to paradox in set theory, we need take notice only of metaphysical kinds represented by *a priori* ideas; for the claim is purely *a priori*.

10.2.2 Preliminaries: the smallest and the greatest set.

We are now in a position to discuss Russell's sophism. We shall approach it slowly. He says that "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\} \equiv \{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$" is logically true, which is the same as saying that "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\} \vee \{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$" is logically false; we shall consider the problem in the latter form. There the proposition is an instance of the principle of excluded middle, and so should be logically true; our objective will be to show that the proposition is in fact logically true, and only true; not both true and false.

We shall begin by musing a while about the predicates "$\xi\notin\xi$" and "$\xi\in\xi$", likening them to the predicates "$\eta=\eta$" and "$\eta\neq\eta$", where "η", "ξ" are variables for any item whatever, not necessarily a set. We may hold that nothing instantiates the property represented by "$\eta\neq\eta$" — that is, $\lceil\eta\neq\eta\rceil$ — so that the extension of "$\eta\neq\eta$" is the null-set, $\varnothing$; and we may take it that $\varnothing$ is the smallest set. Accordingly, we may hold that everything instantiates $\lceil\eta=\eta\rceil$, so that the extension of "$\eta=\eta$", if there were one, would be the set of everything; call it "○". The trouble with ○ is that if it did not belong to itself, it would not be itself: *i.e.*, it would not be the set of everything, and thus would not have a clear and distinct identity; yet if it did belong to itself, it would not be itself either: it would not have a clear and distinct identity, since no set can instantiate the same property which determines its membership. However, this does not require us to conclude that $\bigcirc$ both does and does not belong to itself; rather, we must conclude that — not having a clear and distinct identity — $\bigcirc$ does not exist; and, in turn, we must take this as showing that *there is no greatest set.* Thus although there is the smallest set, there is no greatest set; and this is analogous to there being the smallest natural number, 1, but no greatest number: for, as with $\bigcirc$, supposing there were the greatest number, it would obviously not have a clear and distinct identity.

Similarly, we may hold that nothing instantiates the property represented by "$\xi \in \xi$" — that is, $\ulcorner \xi \in \xi \urcorner$ — so that the extension of "$\xi \in \xi$" is the smallest set, $\varnothing$; for, once again, supposing some set $\{\omega \mid \Psi\omega\}$ instantiated that property, it would instantiate $\ulcorner \Psi\omega \urcorner$: but then it would not be the set it is, for no set can instantiate the property which fixes its membership. Accordingly, we may hold that every set — in fact, not only sets but *everything* — instantiates $\ulcorner \xi \notin \xi \urcorner$, so the extension of "$\xi \notin \xi$", if there were one, would be $\bigcirc$, the set of everything; but $\bigcirc$ does not have a clear and distinct identity, and hence we must conclude that $\bigcirc$ does not exist, and that there is no greatest set (or, specifically, no greatest set of sets), although there is the smallest set, $\varnothing$. We have it therefore that the predicate "$\xi \notin \xi$" has no extension, and the subject-term "$\{\xi \mid \xi \notin \xi\}$" has no referent; in short, the set $\{\xi \mid \xi \notin \xi\}$ does not exist. But is that any cause for worry? This question brings us back to Frege and Russell, for whom it is indeed a serious problem.

For Frege and Russell, the predicate-term "$\xi \notin \xi$" must have an extension (*viz.*, the set $\{\xi \mid \xi \notin \xi\}$), and the subject-term "$\{\xi \mid \xi \notin \xi\}$" must have a referent (again, the set $\{\xi \mid \xi \notin \xi\}$), since otherwise these terms would be meaningless — 'sham symbols', to paraphrase Frege; and this is because they rely on the fallacious *principle of extension-meaning supervenience*, which forces the existence of the set, and hence the paradox. That principle, as I have argued throughout this book, has been the root of much evil in semantics and logic, and in Analytic Philosophy generally; and has eventually blossomed in such a variety of *fleurs du mal* as the contemporary relevant and paraconsistent logics.

Further, we saw earlier that Frege regarded the predicate "is not a member of itself" as 'unexceptionable', since it is undoubtedly meaningful; however, it is exceptionable, insofar as it can be veridically predicated of *any subject-term whatever.* Compare predicates such as "is either blue or not blue", "is either abstract or concrete", "is either universal or particular", "is identical to itself", *etc.*; any proposition of the form "so-and-so is such-and-such", where "is such-and-such" is one of these predicates, will turn out necessarily true.

Such predicates — "$\xi \notin \xi$" among the others — though meaningful, are exceptionable, in that they can be veridically predicated of any subject; and these exceptionable predicates *cannot* and *need not* have an extension; for on the one hand, the extension would have to be the set of everything,

O, which cannot exist since it cannot have a clear and distinct identity (*i.e.*, there cannot be the greatest set); and, on the other hand, any proposition constructed from such a predicate will be either necessarily true or necessarily false, and will not require the existence of that extension for its evaluation. (That Russell's proposition does not require the existence of the extension of "$\xi \notin \xi$" — *viz.*, $\{\xi \mid \xi \notin \xi\}$ — we shall see anon.)

10.2.3 The truth-condition of the sophism.

The foregoing musing might well be enough for an informal refutation of Russell's sophism. In the rest of this section, we shall tighten up the argument, and show that the proposition "$\{\xi \mid \xi \notin \xi\} \in \{\xi \mid \xi \notin \xi\} \vee \{\xi \mid \xi \notin \xi\} \notin \{\xi \mid \xi \notin \xi\}$" is necessarily true and only true, rather than both true and false, by constructing a *truth-condition* for it.

The truth-condition of "$\{\xi \mid \xi \notin \xi\} \in \{\xi \mid \xi \notin \xi\} \vee \{\xi \mid \xi \notin \xi\} \notin \{\xi \mid \xi \notin \xi\}$", for Frege and Russell, is obtained simply by disquotation:

$$\{\xi \mid \xi \notin \xi\} \in \{\xi \mid \xi \notin \xi\} \vee \{\xi \mid \xi \notin \xi\} \notin \{\xi \mid \xi \notin \xi\};$$

for they, believing the principle of extension-meaning supervenience, take it for granted that the extension of the predicate "$\xi \notin \xi$" is the set $\{\xi \mid \xi \notin \xi\}$, so that the referent of "$\{\xi \mid \xi \notin \xi\}$" is $\{\xi \mid \xi \notin \xi\}$; and hence of course the paradox follows. We, however, rejecting the principle of extension-meaning supervenience, cannot take the existence of an extension of a predicate, or the existence of a referent of a subject-term, for granted; for us, the meaning of "$\xi \notin \xi$" consists in its denoting the nominal property $[\xi \notin \xi]$, and the meaning of "$\{\xi \mid \xi \notin \xi\}$" consists in its denoting the nominal property $[\{\xi \mid \xi \notin \xi\}]$, neither of which guarantees the existence of the extension of "$\xi \notin \xi$" or the referent of "$\{\xi \mid \xi \notin \xi\}$". Finding out whether or not there exists such an extension or referent is not merely a matter of semantics, but of epistemology; in general, for any proposition, finding out what the extensions of its constituent ideas are, and hence what its truth-condition is, is a matter of doing some research, either *a posteriori* (in the case of contingent propositions), or *a priori* (in the case of necessary propositions).

The proposition "$\{\xi \mid \xi \notin \xi\} \in \{\xi \mid \xi \notin \xi\} \vee \{\xi \mid \xi \notin \xi\} \notin \{\xi \mid \xi \notin \xi\}$" is purely *a priori*, since it is constructed solely from *a priori* ideas; and so discovering what, if any, extensions or referents its constituent ideas have is a matter of *a priori* research. But now, we may reason that if the idea "$\{\xi \mid \xi \notin \xi\}$" had a referent, or the idea "$\xi \notin \xi$" had an extension, it would be the set $\{\xi \mid \xi \notin \xi\}$, which could not have a clear and distinct identity; for, firstly, it would be a member of itself *iff* it were not; and, secondly, it would be the set of everything, the greatest set, which likewise would be a member of itself *iff* it were not. So "$\xi \notin \xi$" has no extension, and "$\{\xi \mid \xi \notin \xi\}$" has no referent; and this is sufficient to show that the proposition must be true: for if the set does not exist, then it is surely true that the set is not a member of itself, and it is surely false that the set is a member of itself. This accords, as it should, with the more general proposition that no set can be a member of itself, and with the principle of excluded middle: thus the proposition is necessarily true and only true, not both true and false, precisely because the set $\{\xi|\xi\notin\xi\}$ does not exist.

Although the subject-idea "$\{\xi|\xi\notin\xi\}$" has no referent, it does have a clearly defined extension (*qua* set of referents), namely, the null-set; so we might think of expressing the truth-condition of "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\} \vee \{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$" as "$\varnothing\in\varnothing \vee \varnothing\notin\varnothing$", which is true as required. But this formulation is incorrect; for we are now dealing with the *extension* of "$\{\xi|\xi\notin\xi\}$", which is $\varnothing$, not the *referent* of "$\{\xi|\xi\notin\xi\}$", which does not exist; and hence the fact represented by "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\} \vee \{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$" cannot be that $\varnothing\in\varnothing \vee \varnothing\notin\varnothing$, contrary to first impressions. Properly, the truth-condition should be constructed thus: we begin with the semantically simple predicate-idea "$\xi\in\xi$", and then form the complex idea "$\xi\notin\xi$" (that is, "$\neg\xi\in\xi$"); next, we enquire about the extension of "$\xi\notin\xi$", which — if it existed — would be the set $\{\xi|\xi\notin\xi\}$; however, having rejected the principle of extension-meaning supervenience, we cannot take the existence of $\{\xi|\xi\notin\xi\}$ for granted just because "$\xi\notin\xi$" is meaningful; so, in forming the complex subject-idea "$\{\xi|\xi\notin\xi\}$" and the proposition "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\}$", we must, so to speak, raise the particular question about the *referent* of "$\{\xi|\xi\notin\xi\}$", hence going *beyond the meaning* of the idea and the proposition, to metaphysical reality as such. The truth-condition of "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\}$" — *i.e.*, the *sufficient and necessary condition for the truth of* "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\}$" — may be then formulated as:

$$\textit{(i)}\qquad (\Sigma\zeta)(\zeta=\{\xi|\xi\notin\xi\} \wedge \zeta\in\zeta);$$

in other words, "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\}$" is true *iff* $(\Sigma\zeta)(\zeta=\{\xi|\xi\notin\xi\} \wedge \zeta\in\zeta)$.

Turning now to the other disjunct, *viz.*, "$\{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$", we may analogously construct the truth-condition $(\Sigma\zeta)(\zeta=\{\xi|\xi\notin\xi\} \wedge \zeta\notin\zeta)$, noting that although *sufficient*, it is *not necessary*; for $\neg(\Sigma\zeta)\zeta=\{\xi|\xi\notin\xi\}$ would also ensure the truth of "$\{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$". Here we come to the crux of the matter: if the condition $(\Sigma\zeta)(\zeta=\{\xi|\xi\notin\xi\} \wedge \zeta\notin\zeta)$ obtained, "$\{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$" would be true; but if the condition $\neg(\Sigma\zeta)\zeta=\{\xi|\xi\notin\xi\}$ obtained, "$\{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$" would likewise be true; so the sufficient *and necessary* condition is:

$$\textit{(ii)}\qquad (\Sigma\zeta)(\zeta=\{\xi|\xi\notin\xi\} \wedge \zeta\notin\zeta) \vee \neg(\Sigma\zeta)\zeta=\{\xi|\xi\notin\xi\}.$$

Combining *(i)* and *(ii)*, the truth-condition of Russell's sophistical proposition becomes:

$$\textit{(iii)}\ (\Sigma\zeta)(\zeta=\{\xi|\xi\notin\xi\} \wedge (\zeta\in\zeta \vee \zeta\notin\zeta)) \vee \neg(\Sigma\zeta)\zeta=\{\xi|\xi\notin\xi\}.$$

In summary, "$\{\xi|\xi\notin\xi\}\in\{\xi|\xi\notin\xi\} \vee \{\xi|\xi\notin\xi\}\notin\{\xi|\xi\notin\xi\}$" is true *iff either* there is the set $\{\xi|\xi\notin\xi\}$ and it either is or is not a member of itself, *or* there is no such set; the former leg, when taken on its own, leads to the paradox; the latter, on its own or in disjunction, resolves the paradox: Russell's proposition is therefore true and only true, not both true and false.

We may note how the truth-condition *(iii)* links to the two horns of Frege's dilemma, reviewed in 10.1: that either one must put up with *paraconsistency*, or else one must regard the apparently meaningful predicate-term "$\xi \notin \xi$" as having *no extension*, and the apparently meaningful subject-term "$\{\xi \mid \xi \notin \xi\}$" as having *no referent*; the latter alternative seemed to Frege impossible because of his adherence to the rule of extension-meaning supervenience, which lay in the foundation of his set theory; the former, which to Frege would have been the end of logic, many a contemporary hell's logician has built his name upon. But neither party has considered the certainly correct alternative that meaning does not fix extension, that the meaningfulness of a predicate or set-name does not guarantee the existence of the set, and that set theory cannot take off the ground without a sound semantical theory in its foundation; hence so much confusion in modern logic and Analytic Philosophy.

The truth-condition *(iii)* — more precisely, the statement of it — should not be thought of as giving the *meaning* of Russell's sophistical proposition; for one thing, there is nothing sophistical about that statement: it shows the proposition clearly and unequivocally true, the left disjunct being logically false whilst the right true; rather, we should think of the statement as specifying *the sufficient and necessary condition for the truth of the proposition*, and therefore reaching (by *a priori* means alone) beyond meaning to intelligible reality as such.

Comparing this truth-condition with that we find implicit in Russell and Frege, which is got simply by disquotation — that is,

$$\{\xi \mid \xi \notin \xi\} \in \{\xi \mid \xi \notin \xi\} \ \vee \ \{\xi \mid \xi \notin \xi\} \notin \{\xi \mid \xi \notin \xi\} \ –$$

it is obvious that the fallacy behind the sophism consists in supposing the existence of the set $\{\xi \mid \xi \notin \xi\}$ merely on the grounds that the predicate-term "$\xi \notin \xi$" and the subject-term "$\{\xi \mid \xi \notin \xi\}$" are meaningful; but deeper, underlying the fallacy, there is an inability to distinguish consistently between a symbol and a thing symbolised. (Russell was notorious not only for the 'paradox' and for Analytic Philosophy, but also for being unable to distinguish between the snowfields of Mont Blanc and propositions about Mont Blanc, not to mention such subtle distinctions as that between a symbol and "symbol"; so perhaps the 'paradox' was no accident ...)

The lesson is, as it has been throughout much of this book, that the meaning of a term does not determine its extension (or set of referents), and the meaning of a proposition does not determine its truth-condition; the extensional-*cum*-truth-conditional theory of meaning — in general, the semantic principle of extension-meaning supervenience — is false, not only as a matter of fact but *necessarily* so; so a set theory, *qua* extension theory, cannot be founded on that principle, on pain of containing paradoxical propositions. The existence of the extension or set of referents, and hence the identity of the truth-condition, cannot be taken for granted merely because the term and the proposition are meaningful; they must be *discovered*, either by *a priori* research (in the case of demonstrative propositions), or by *a posteriori* research (in the case of contingent propositions). Russell's sophism, and Frege's set theory, rest on taking it for granted that meaning determines extension and truth-condition; hence the paradox. Once the old sorcerer's metaphysic of extension-meaning supervenience is rejected, it becomes obvious, with a bit of *a priori* rumination, that the *sufficient and necessary condition* for the truth of "$\{\xi \mid \xi \notin \xi\} \in \{\xi \mid \xi \notin \xi\} \vee \{\xi \mid \xi \notin \xi\} \notin \{\xi \mid \xi \notin \xi\}$" is not that there is the set $\{\xi \mid \xi \notin \xi\}$ and it either is or is not a member of itself; for that condition is sufficient but not necessary; there being no such set is also sufficient: so the sufficient and necessary condition must be a disjunction of the two.

There are, of course, many variants of the sophism; the well-known barber version differs from the set version only in that it is built upon the predicate "does not shave himself", which is an *a posteriori* predicate, in contrast to the *a priori* predicate used in the set version. Nor is there anything special about the art of barbery; other professions will do equally well. "The cook who cooks for all and only people who do not cook for themselves" raises the same conundrum. Alas, that marvellous fellow does not, nor can exist; yet, if it be any comfort, we may still *meaningfully* dream about him, and the proposition that he either does or does not cook for himself is still true, and only true.

10.3 The Well-Formedness of the Mind

The mind, according to CTM, comprises a basis of semantically simple *a posteriori* ideas, a class of generative operations — laden with *a priori* ideas — for the production of complex ideas, a class of generative operations — also laden with *a priori* ideas — for the production of propositions from the simple and complex ideas, and a class of propositional cognitive and emotive operations, likewise laden with *a priori* ideas. The fact that any proposition, as *token* of a mental sentence, is generable from the simple *a posteriori* and *a priori* ideas (with some propositions, from the simple *a priori* ideas alone) ensures — I think we may safely conjecture — that the proposition will have a uniquely fixed and discoverable logical modality, either necessary truth, falsehood, or contingency; for the proposition, being thus generable from its simple constituent ideas, must be accordingly *synthesisable* — and, conversely, *analysable* — from the semantic identity of the simple ideas, and so provable either as necessarily true, or as false, or as contingent. Hence we may conjecture that the human mind is epistemically *well-formed*, in that any proposition it can contemplate has a unique and discoverable logical modality, and where the modality is necessary truth or falsehood, the mind can know the truth or falsehood with apodeictic certainty; there are no propositions both true and false, nor any neither true nor false (let alone any bearing some other invented truth-values).

Frege's project of a logical *begriffsschrift* lay in the origin of Analytic Philosophy, and defined many of its characteristic features: its anti-mentalism, its formal semantics based on the set theory of extension-meaning supervenience, its paradoxes, its confused search for the correct notion of logical implication and deductive validity, its 'results' concerning logical decidability, and so forth. The project as such is, without question, worthwhile and profound; but it cannot succeed unless the logical *begriffsschrift* is construed as a model of the symbolic system of the mind; and only when it is so construed, the project may truly become a part of the classical philosophical tradition dating back at least to Plato.

Epilogue

I would like to say a few words about what has been done in the Classical Theory of Mind, what is yet to be done, how it is to be done, and what we may expect to gain by doing it. CTM has had a long history, with many ups and downs. Of late, it flourished in the 17th century, struggled for survival in the 18th, and very nearly disappeared after Kant. One might wonder why it has fallen into oblivion. An answer could be that it relied on introspection as its sole method of enquiry, and introspection became unsuitable for psychology as a natural science. There are problems with this answer. According to CTM, psychology is — in part, though not entirely — a demonstrative, metaphysical science, which may not be fully naturalisable; the trouble with turning CTM into a natural science should not have been a decisive reason for abandoning it. But perhaps the reason was that people were no longer able to make sense of it, because of accumulated mistakes by CTM theorists. I rather favour this explanation. Here are some errors which would have served to obfuscate the project of CTM.

Kant construed conceptual analysis and synthesis as mutually exclusive and incompatible, so that a true *a priori* analytic proposition could not be *a priori* synthetic, and conversely. What confusion this has subsequently occasioned is difficult to over-estimate; witness only the wranglings in Analytic Philosophy about *a priori* synthetic knowledge. Again, he construed *a priori* concepts as functioning to organise *a posteriori* concepts into the manifold of intuition, but without veridically representing the world as it is; so that the natural world — the world physical sciences aim to describe — turned out to be the human mind's nominal or phenomenal world, rather than a world independent from the mind. Rightly so, few were willing to regard nature as a figment of the imagination, ordered though it be by the *a priori* faculty.

Locke failed to distinguish explicitly between empirical (*a posteriori*) and non-empirical (*a priori*) ideas, or concepts, taking all ideas as empirical. An attentive reader of the *Essay* will find the distinction implicitly there; nevertheless, Locke's failure to draw explicitly the distinction led to the extreme empiricism of Berkeley and Hume, which Locke himself would not have endorsed; and eventually it led to the disintegration of the CTM framework, already evident in Hume.

Descartes allowed himself to think of material substance as consisting in extension alone; and finding the mind *really distinct* from the body (so that God, if nothing else, could separate the two), he concluded that the mind is unextended. Yet since the mind and the body are in causal interaction, he was compelled to believe that the unextended causally interacts with the extended. How much damage this has done to the project of CTM I need not tell.

Lastly, Plato (*e.g.*, 596–597) held that for *any* meaningful predicate, there is a *real* universal, a mind-independent and *non-inherent* (as it were, free-floating) form which the predicate represents, and which all particular objects the predicate refers to share in common. Quine has called this view "Plato's beard", connoting I suppose its venerable influence on subsequent philosophy. Along with Wittgenstein, and others on the Radical-*cum*-Middlebrow side of Analytic Philosophy, he recommended *behavioural holism* to correct Plato's mistake — not unlike a barber who would shave Plato's beard by cutting his head off. (Evil tongues insinuate that this was not Plato's gravest error; the worst was to have founded the Academia; for no sooner had he died than it turned, so to speak, *Speusippian*, and it has been so more or less ever since. Perhaps, then, it serves him right to have met with an Academic barber in the Academia who shaves all and only Academics who do not shave themselves.)

Whether Plato's gravest error could ever be rectified I dare not predict. But as to the other mistakes of CTM theorists, these can be rectified, and CTM need not fall with them. The project of CTM, in general, is to work out the syntax, semantics, and the epistemology of the mind.

Concerning the *syntax*, the fundamental task is to map out the mental code in its *a priori* and *a posteriori* portions. This is a very ambitious undertaking. Until and including Descartes, CTM theorists assumed that there is a code of simple ideas, whereof complex ideas and propositions are made, but without attempting to spell out just what it is. Locke was the first to take up the challenge of describing the mind, as a system of ideas, in detail. He recognised that it was not practically possible, by his introspective method and with the public languages at his disposal, to list each type of simple idea the mind can form; for one thing, ordinary public languages have words for categories of simple ideas, not for each type of simple idea (*e.g.*, the word "sweet" stands for a whole range of simple ideas **sweet**, which we distinguish introspectively but not publicly); for another, the types of simple idea are too numerous and finely grained to be listed by any one researcher. Still, Locke's account of the mental code remains unsurpassed to this day.

Concerning the *semantics* of CTM, I set out its primary notions in Chapters 7 and 9: that the meaning of an idea, simple or complex, consists in that the idea denotes a certain property, and that the identity of the property is fixed nominally, by the *form of consciousness* associated with each tokening of the idea; for example, the idea **blue** means what it does since it denotes the nominal property [blue], and the identity of [blue] is determined by the form of consciousness ⟦blue⟧ essentially associated with each tokening of **blue**. The problem of what constitutes such forms of consciousness as ⟦blue⟧ of the idea **blue** is the most difficult in CTM. I emphasised that an answer to the question of the *natural implementation* of an idea such as **blue** is not an answer, nor does it constrain one's answer, to the question of what constitutes the form of consciousness ⟦blue⟧. I put forward a genetic account of the natural implementation of ideas, but without committing myself to materialism about their forms of consciousness; nor am I committed to dualism, or 'neutral monism', *etc*. These standard distinctions will not be good enough to deal with the problem.

Concerning the *epistemology* of CTM, the project is to account for the analytic, synthetic, and deductive cognitive operations and processes. Analysis is the purest form of confirmation; however, from the point of view of the ontogeny of human understanding, synthesis is the more fundamental. The infant mind begins from the simple empirical ideas, and proceeds to construct complex ideas, propositions and structures of propositions, guiding their synthesis by several kinds of evidence: observational evidence, relying on the simple empirical ideas (together with the *a priori* ideas) comprising the proposition being synthesised; holistic evidence, relying on other propositions already established to some degree of likelihood (in addition to the proposition's constituent empirical and *a priori* ideas); eventually, the mind becomes able to synthesise a proposition solely on the evidence of its constituent *a priori* ideas, thereby coming to know the proposition with *a priori* certainty; and finally, the mind may become able to prove the same or other propositions solely by *a priori* analysis.

Introspection was traditionally the only means available to researchers working on the syntax, semantics, and epistemology of the mind; and they achieved a great excellence in it. The reason why introspection fell into disrepute as a method of psychology in the 19th and the present centuries was not so much that knowledge of the mind cannot be gained by it (in fact, knowledge of the mind is gained *first and foremost* by it), but that the art of it had dissipated, and that it was often applied inappropriately. CTM's position is that introspective *universality* — from person to person, and from time to time — is to be expected at the level of the mind's foundation of *simple* empirical and non-empirical ideas, and operations on ideas; certainly not at the level of complex ideas, propositions, and complex operations, which more or less vary from person to person and time to time. Accordingly, introspection is appropriately applied at the level of that universal foundation, and even there it requires much skill and attention; when it comes to the personal psychology of complex ideas and propositions, we may and do employ introspection for our own purposes, but it would be a mistake to expect universality; and it would be naïve to expect reliability in introspective reports from individuals who are not skilled in and accustomed to it. For all its shortcomings, though, introspection will remain the first way of psychology, however the moderns may dislike it;

without introspection, there would be no psychology to speak of, no minds to enquire into.

In addition to introspection, the project of mapping out the structure and function of the mental code may now be able to deploy methods which were not traditionally available. *Firstly*, the project could be aided by the use of *formal languages* and models of the mind. The classical theorists did not have formal languages at their disposal, and had to rely entirely upon ordinary natural languages; perhaps much of the seeming obscurity in their accounts, and disagreements between them, were due to the vagueness and relative clumsiness of natural languages. Formal coding has much the same advantages in CTM as in logic and mathematics, but also the same dangers. The proliferation of formal systems in present-day axiomatic logic, natural deduction, and the so-called "possible-worlds semantics", has left most philosophers in the Analytic movement with the impression that logic, along with other sciences, is but a matter of convention and pragmatic use; systems of logic are made or unmade depending on their use in a larger context, without universal validity. A similar proliferation of formal models in CTM could easily lead to the like conclusion about the structure and function of the mind; to avoid it, one must regard formal modelling merely as a means toward the end of describing the mind as such, rather than as a linguistic end in itself. *Secondly*, the project of mapping out the mental code could be aided by the natural science of the implementation of the mind in the brain, which was likewise unavailable to the classical theorists. I have proposed that an ideational system such as that required by CTM could not be implemented in the brain by anything other than the genes; and that the genes are eminently suited for that function, as regards the implementation of ideas, their acquisition, memory, recollection, and rational and associative processing.

The project of mapping out the human genome has been under way for some time; I would not be surprised if it turned out to include — as an essential, species-determining part — the mapping out of the human system of semantically simple *a posteriori* and *a priori* ideas, and operations on ideas.

Bibliography

Ackermann, W. (1956), 'Begründung einer strengen Implikation', *The Journal of Symbolic Logic, 21,* pp. 113–128.

Agranoff, B.W., Burrell, H.R., Dokas, L.A., & Springer, A.D. (1978), 'Progress in Biochemical Approaches to Learning and Memory'. In M.A. Lipton, A. DiMascio, & K.F. Killam (eds.), *Psychopharmacology: A Generation of Progress* (pp. 623–635). New York: Raven Press.

Agranoff, B.W., Davis, R.E., & Brink, J.J. (1965), 'Memory Fixation in the Goldfish', *Proceedings of National Academy of Sciences USA, 54,* pp. 788–793.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J.D. (1989), *Molecular Biology of the Cell,* 2nd ed. (chap. 19). New York: Garland.

Anderson, A.R., Belnap, N.D., Jr., *et al.* (1975), *Entailment: the logic of relevance and necessity,* Volume 1. Princeton: Princeton University Press.

Armstrong, R.C., & Montminy, M.R. (1993), 'Transsynaptic Control of Gene Expression', *Annual Review of Neuroscience, 16,* pp. 17–29.

Bailey, C.H., & Kandel, E.R. (1993), 'Structural Changes Accompanying Memory Storage', *Annual Review of Physiology, 55,* pp. 397–426.

Bailey, C.H., Hawkins, R.D., & Chen, M. (1983), 'Uptake of [^{3}H] Serotonin in the Abdominal Ganglion of *Aplysia Californica*: Further Studies on the Morphological and Biochemical Basis of Presynaptic Facilitation', *Brain Research, 272,* pp. 71–81.

Bliss T.P.V., & Lømo, T. (1973), 'Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anaesthetized Rabbit Following Stimulation of the Perforant Path', *The Journal of Physiology (London), 232,* pp. 331–356.

Bliss, T.V.P., & Lynch, M.A. (1988), 'Long-term Potentiation of Synaptic Transmission in the Hippocampus: Properties and Mechanisms'. In P.W. Landfield & S.A. Deadwyler (eds.), *Long-term Potentiation: From Biophysics to Behavior* (pp. 3–72). New York: Liss.

Burge, T. (1979), 'Individualism and the Mental'. In *Midwest Studies in Philosophy*, Vol. IV: *Studies in Metaphysics*, P.A. French, T.E. Uehling, H.K. Wettstein, (eds.). Minneapolis, Minn.: University of Minnesota Press.

Byrne, W.L., Samuel, D., Bennett, E.L., Rosenzweig, M.R., Wasserman, E., Wagner, A.R., Gardner, F., Galambos, R., Berger, B.D., Margules, D.L., Fenichel, R.L., Stein, L., Corson, J.A., Enesco, H.E., Chorover, S.L., Holt, C.E., Schiller, P.H., Chippetta, L., Jarvik, M.E., Leaf, R.C., Dutcher, J.D., Horovitz, Z.P., & Carlson, P.L. (1966), 'Memory Transfer', *Science, 153*, pp. 658-9.

Cajal, S.R.y. (1988/1911), *Histologie du systèm nerveux de l'homme et des vertébrés*, vol. 2 (Paris: Maloine). Excerpts translated by J. DeFelipe & E. Jones, *Cajal on the Cerebral Cortex*. New York: Oxford University Press.

Cajal, S.R.y. (1990/1894), *New Ideas on the Structure of the Nervous System in Man and Vertebrates*. Translated by N. Swanson & L.W. Swanson. Originally published as *Les nouvelles idées sur la structure du systèm nerveux chez l'homme et chez les vertebrés* (Paris: Reinwald & Cie). Cambridge, Mass.: MIT Press.

Carnap, R. (1932–1933), 'Psychology in Physical Language'. In *Logical Positivism*, Alfred J. Ayer (ed.). New York: The Free Press, 1959. First published: 'Psychologie in physikalischer Sprache', *Erkenntnis*, 1932–1933, Vol. III.

Carnap, R. (1936–1937), 'Testability and Meaning', *Philosophy of Science*, 1936, Vol. 3, No. 4, pp. 419–471; 1937, Vol. 4, No. 1, pp. 1–40.

Carnap, R. (1928), *Der logische Aufbau der Welt*. Berlin: Weltkreis-Verlag.

Chomsky, N. (1986), *Knowledge of Language: its nature, origins, and use*. New York: Praeger.

Cole, A.J., Saffen, D.W., Baraban, J.M., & Worley, P.F. (1989), 'Rapid increase of an immediate early gene messenger RNA in hippocampal neurons by synaptic NMDA receptor activation', *Nature, 340*, pp. 474–476.

Couture, J., & Nielsen, K. (1993), 'On Construing Philosophy', *The Canadian Journal of Philosophy: Supplementary Volume 19*, pp. 1–55.

Crick, F.H.C., & Asanuma, C. (1986), 'Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex'. In J.L. McClelland, D.E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models* (pp. 333–371). Cambridge, Mass.: MIT Press.

Crick, F.H.C. (1963), 'The Recent Excitement in the Coding Problem', *Progress in Nucleic Acid Research, 1*, pp. 164–217.

Descartes, R. (1984), 'Meditations on First Philosophy'. In *The Philosophical Writings of Descartes*, translated and edited by Cottingham, J., Stoothoff, R., Murdoch, D. Cambridge: Cambridge University Press.

Dudai, Y. (1989), *The Neurobiology of Memory: Concepts, Findings, Trends*. Oxford: Oxford University Press.

Dunn, A.J. (1986), 'Biochemical Correlates of Learning and Memory'. In J.L. Martinez, Jr., & R.P. Kesner (eds.), *Learning and Memory: A Biological View* (pp. 165–201). London: Academic Press.

Dunn, A.J. (1976), 'The Chemistry of Learning and the Formation of Memory'. In W.H. Gispen (ed.), *Molecular and Functional Neurobiology* (pp. 347–387). Amsterdam: Elsevier.

Flexner, L.B., Flexner, J.B., & Roberts, R.B. (1967), 'Memory in Mice Analyzed with Antibiotics', *Science, 155*, pp. 1377–1383.

Fodor, J.A. (1975), *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J.A. (1981), 'The Present Status of the Innateness Controversy'. In *Representations: Philosophical Essays on the Foundations of Cognitive Science* (pp. 257–316). Cambridge, Mass.: MIT Press.

Fodor, J.A. (1982), 'Cognitive Science and the Twin-Earth Problem', *Notre Dame Journal of Formal Logic*, Vol. 23, No. 2, pp. 98–118.

Fodor, J.A. (1986), 'Banish disContent'. In *Language, Mind and Logic*, Jeremy Butterfield (ed.). Cambridge: Cambridge University Press.

Fodor, J.A. (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.

Fodor, J.A. (1990a), 'A Theory of Content, I: The Problem', *A Theory of Content and Other Essays*. Cambridge, Mass.: The MIT Press.

Fodor, J.A. (1990b), 'A Theory of Content, II: The Theory', *A Theory of Content and Other Essays*. Cambridge, Mass.: The MIT Press.

Fodor, J.A., & Pylyshyn, Z.W. (1988), 'Connectionism and cognitive architecture: A critical analysis', *Cognition, 28*, 3–71.

Fodor, J.A. (1994), *The Elm and the Expert: Mentalese and Its Semantics*. Cambridge, Mass.: The MIT Press.

Freeman, W.J., & Skarda, C.A. (1990), 'Representations: Who Needs Them?'. In J.L. McGaugh, N.M. Weinberger, & G. Lynch (eds.), *Brain Organization and Memory: Cells, Systems, and Circuits* (pp. 375–380). New York: Oxford University Press.

Frege, G. (1884), *The Foundations of Arithmetic*. Oxford: Basil Blackwell, 1950. Translation by J. Austin of *Die Grundlagen der Arithmetik*. Breslau: Verlag von Wilhelm Koebner, 1884.

Frege, G. (1892), 'On Sense and Reference', *Translations from the Philosophical Writings of Gottlob Frege*, Peter Geach and Max Black, eds. Oxford: Basil Blackwell, 1952; second edition (with corrections), 1960. Translation by Max Black of 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik*, 1892, Vol. 100, pp. 25–50.

Frege, G. (1893), *Grundgesetze der Arithmetic, begriffsschriftlich abgeleitet*, Band I. Jena: Verlag Herman Pohle, 1893. Partial translation by Montgomery Furth, ed.: *Gottlob Frege, The Basic Laws of Arithmetic*. Berkeley: University of California Press, 1964. (Also partially translated in P. Geach and M. Black, eds.: *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell, 1952; second edition (with corrections), 1960.)

Frege, G. (1903), *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, Band II. Jena: Verlag Herman Pohle, 1903. Partial translation in P. Geach and M. Black, eds.: *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell, 1952; second edition (with corrections), 1960. (Also partially translated by Montgomery Furth, ed.: *Gottlob Frege, The Basic Laws of Arithmetic*. Berkeley: University of California Press, 1964.)

Frege, G. (1980), *Philosophical and Mathematical Correspondence*, edited by Gabriel, G., Hermes, H., Kambartel, F., Thiel, C., Veraart, A., abridged for the English edition by McGuinness, B., and translated by Kaal, H. Oxford: Basil Blackwell.

Frey, U., Krug, M., Reymann, K.G., & Matthies, H. (1988), 'Anisomycin, an inhibitor of protein synthesis, blocks late phases of LTP phenomena in the hippocampal CA_1 region *in vitro*', *Brain Research, 452*, pp. 57–65.

Geach, P.T. (1958), 'Entailment', *The Aristotelian Society: Supplementary Volume* XXXII, pp. 157–172.

Goelet, P., Castellucci, V.F., Schacher, S., & Kandel, E.R. (1986), 'The long and the short of long-term memory — a molecular framework', *Nature, 322*, pp. 419–422.

Goldman-Rakic, P.S. (1987), 'Circuitry of primate prefrontal cortex and regulation of behavior by representational memory'. In Geiger, S.R. (Executive ed.), Mountcastle, V.B. (Section ed.), & Plum, F. (Volume ed.), *Handbook of Physiology*, Section 1, Volume V: *Higher Functions of the Brain, Part 1* (pp. 373–417). American Physiological Society: Bethesda, Maryland.

Greenberg, S.M., Castellucci, V.F., Bayley, H., & Schwartz, J.H. (1987), 'A molecular mechanism for long-term sensitization in *Aplysia*', *Nature, 329*, pp. 62–65.

Haley, J.E., Wilcox, G.L., & Chapman, P.F. (1992), 'The role of nitric oxide in hippocampal long-term potentiation', *Neuron, 8*, pp. 211–16.

Hawkins, R.D., Kandel, E.R., & Siegelbaum, S.A. (1993), 'Learning to Modulate Transmitter Release: Themes and Variations in Synaptic Plasticity', *Annual Review of Neuroscience, 16*, pp. 625–65.

Hume, D. (1975), 'An Enquiry Concerning Human Understanding'. In L.A. Selby-Bigge & P.H. Nidditch (eds.), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press.

Hume, D. (1978), *A Treatise of Human Nature*. Edited by L.A. Selby-Bigge & P.H. Nidditch. Oxford: Oxford University Press.

Hydén, H., & Egyházi, E. (1964), 'Changes in RNA Content and Base Composition in Cortical Neurons of Rats in a Learning Experiment Involving Transfer of Handedness', *Proceedings of National Academy of Sciences USA, 52*, pp. 1030–1035.

Jacobson, A.L., Babich, F.R., Bubash, S., & Jacobson, A. (1965), 'Differential Approach Tendencies Produced by Injection of RNA from Trained Rats', *Science, 150*, pp. 636–637.

Johnston, D., Williams, S., Jaffe, D., & Gray, R. (1992), 'NMDA-Receptor-Independent Long-Term Potentiation', *Annual Review of Physiology, 54*, pp. 489–505.

Kandel, E.R., & Tauc, L. (1965), 'Heterosynaptic facilitation in neurones of the abdominal ganglion of *Aplysia depilans*', *The Journal of Physiology (London), 181*, pp. 1–27.

Kandel, E.R. (1991), 'Cellular Mechanisms of Learning and the Biological Basis of Individuality'. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (eds.), *Principles of Neural Science*, 3rd ed. (pp. 1009–1031). London: Prentice Hall International.

Kandel, E.R. (1989), 'Genes, Nerve Cells, and the Remembrance of Things Past', *Journal of Neuropsychiatry and Clinical Neurosciences, 1*, pp. 103–125.

Kant, I. (1977), 'Prolegomena to Any Future Metaphysics'. In S.M. Cahn (ed.), *Classics of Western Philosophy* (pp. 933–1008). Indianapolis: Hackett Publishing Company. Paul Carus translation, revised by James W. Ellington.

Klann, E., & Sweatt, J.D. (1990), 'Persistent alteration of protein kinase activity during the maintenance phase of long-term potentiation', *Society for Neuroscience Abstracts, 16*, p. 144.

Lashley, K.S. (1950), 'In Search of the Engram', *Symposium of the Society for Experimental Biology, 4*, pp. 454–482. New York: Cambridge University Press.

Leibniz, G. W. (1981), *New Essays on Human Understanding*, translated and edited by Remnant, P., and Bennett, J. Cambridge: Cambridge University Press.

Locke, J. (1975), *An Essay Concerning Human Understanding*, edited by Nidditch, P. H. New York: Oxford University Press.

Madison, D.V., Malenka, R.C., & Nicoll, R.A. (1991), 'Mechanisms Underlying Long-Term Potentiation of Synaptic Transmission', *Annual Review of Neuroscience, 14*, pp. 379–97.

Malinow, R., Madison, D.V., & Tsien, R. (1988), 'Persistent protein kinase activity underlying long-term potentiation', *Nature, 335*, pp. 820–24.

McConnell, J.V. (1962), 'Memory Transfer Through Cannibalism in Planarians', *Journal of Neuropsychiatry (Suppl. 1), 3*, pp. 42–48.

McConnell, J.V. (1971), 'Confessions of a scientific humorist'. In J.V. McConnell, & M. Schutjer (eds.), *Science, sex, and sacred cows*. New York: Harcourt Brace Jovanovich.

McNaughton, B.L. (1982), 'Long-term synaptic enhancement and short-term potentiation in rat fascia dentata act through different mechanisms', *The Journal of Physiology (London), 324*, pp. 249–62.

Milner, B. (1966), 'Amnesia following operations on the temporal lobes'. In C.W.M. Whitty & O.L. Zangwill (eds.), *Amnesia* (pp. 109–33). London: Butterworths.

Montarolo, P.G., Goelet, P., Castellucci, V.F., Morgan, J., Kandel, E.R., & Schacher, S. (1986), 'A Critical Period for Macromolecular Synthesis in Long-Term Heterosynaptic Facilitation in *Aplysia*', *Science, 234*, pp. 1249–1254.

Montminy, M.R., & Bilezikjian, L.M. (1987), 'Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene', *Nature, 328*, pp. 175–178.

Morgan, J.I., & Curran, T. (1991), 'Stimulus-Transcription Coupling in the Nervous System: Involvement of the Inducible Proto-oncogenes *fos* and *jun*', *Annual Review of Neuroscience, 14*, pp. 421–51.

Penfield, W., and Roberts L. (1959), *Speech and Brain Mechanisms*. Princeton: Princeton University Press.

Pinsker, H.M., Hening, W.A., Carew, T.J., & Kandel, E.R. (1973), 'Long-term Sensitisation of a Defensive Withdrawal Reflex in *Aplysia*', *Science, 182*, pp. 1039–1042.

Putnam, H. (1975), 'The Meaning of "Meaning"', *Mind, Language and Reality*. Cambridge: Cambridge University Press.

Putnam, H. (1978a), 'Meaning and Knowledge', *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

Putnam, H. (1978b), 'Reference and Understanding', *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

Putnam, H. (1978c), 'Realism and Reason', *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

Putnam, H. (1981), *Reason, Truth and History*. Cambridge: Cambridge University Press.

Putnam, H. (1988), *Representation and Reality*. Cambridge, Mass.: The MIT Press.

Quine, W.V.O. (1948), 'On What There Is', *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press, 1953. First published: *Review of Metaphysics*, 1948, Vol. 2, pp. 21–38.

Quine, W.V.O. (1951), 'Two Dogmas of Empiricism', *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press, 1953. First published: *Philosophical Review*, 1951, Vol. LX, pp. 20–43.

Quine, W.V.O. (1953a), 'The Problem of Meaning in Linguistics', *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.

Quine, W.V.O. (1953b), 'Logic and the Reification of Universals', *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.

Quine, W.V.O. (1953c), 'Notes on the Theory of Reference', *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.

Quine, W.V.O. (1953d), 'Reference and Modality', *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.

Quine, W.V.O. (1960), *Word and Object*. Cambridge, Mass.: The MIT Press.

Quine, W.V.O. (1966a), 'On Mental Entities', *The Ways of Paradox and Other Essays*. New York: Random House.

Quine, W.V.O. (1966b), 'Necessary Truth', *The Ways of Paradox and Other Essays*. New York: Random House.

Quine, W.V.O. (1969), *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (1969a), 'Speaking of Objects', *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (1969b), 'Ontological Relativity', *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (1969c), 'Natural Kinds', *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (1969d), 'Epistemology Naturalized', *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (1970), 'Philosophical Progress in Language Theory', *Metaphilosophy*, Vol. 1, No. 1, pp. 2–19.

Quine, W.V.O. (1987), 'Indeterminacy of Translation Again', *The Journal of Philosophy*, Vol. LXXXIV, No. 1, pp. 5–10.

Richards (1978), *The Language of Reason*. Rushcutters Bay: Pergamon Press (Australia).

Rumelhart, D.E., & McClelland, J.L. (1986), 'PDP Models and General Issues in Cognitive Science'. In D.E. Rumelhart, J.L. McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. (pp. 110–146). Cambridge, Mass.: MIT Press.

Russell, B. (1918–1919), 'The Philosophy of Logical Atomism', *Logic and Knowledge*, Robert C. Marsh (ed.). London: George Allen & Unwin, 1956. First published: *The Monist*, 1918–1919.

Russell, B. (1905), 'On Denoting', *Logic and Knowledge*, Robert C. Marsh (ed.). London: George Allen & Unwin, 1956. First published: *Mind*, 1905, Vol. XIV, pp. 479-493.

Russell, B. (1956), *Logic and Knowledge (Essays 1901–1950)*, edited by Robert C. Marsh. London: George Allen & Unwin.

Schuman, E.M., & Madison, D.V. (1991), 'A Requirement for the Intracellular Messenger Nitric Oxide in Long-Term Potentiation', *Science, 254*, pp. 1503-6.

Schwartz, J.H., & Greenberg, S.M. (1987), 'Molecular Mechanisms for Memory: Second-Messenger Induced Modifications of Protein Kinases in Nerve Cells', *Annual Review of Neuroscience, 10*, pp. 459–76.

Searle, J. (1983), *Intentionality: An essay in the philosophy of mind.* Cambridge: Cambridge University Press.

Shuster, M.J., Camardo, J.S., Siegelbaum, S.A., & Kandel, E.R. (1985), 'Cyclic AMP-dependent protein kinase closes the serotoninsensitive K^+ channels of *Aplysia* sensory neurones in cell-free membrane patches', *Nature, 313*, pp. 392–95.

Siegelbaum, S.A., Camardo, J.S., & Kandel, E.R. (1982), 'Serotonin and cAMP close single K^+ channels in *Aplysia* sensory neurones', *Nature, 299*, pp. 413–17.

Smiley, T.J. (1959), 'Entailment and Deducibility', *Proceedings of the Aristotelian Society*, Vol. LIX, pp. 233-254.

Smolensky, P. (1988), 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences*, 1988, No. 11, pp. 1–74.

Squire, L.R. (1992), 'Memory and the Hippocampus: A Synthesis from Findings with Rats, Monkeys and Humans', *Psychological Review, 99*, pp. 195–231.

Strawson, P.F. (1948), 'Necessary Propositions and Entailment-Statements', *Mind*, Vol. LVII, pp. 184–200.

Thompson, R.F., & Donegan, N.H. (1986), 'The Search for the Engram'. In J.L. Martinez, Jr., & R.P. Kesner (eds.), *Learning and Memory: A Biological View* (pp. 3–52). London: Academic Press.

Von Wright, G.H. (1957), 'The Concept of Entailment', *Logical Studies*. London: Routledge and Kegan Paul, pp. 166–191.

Watling, J. (1958), 'Entailment', *The Aristotelian Society: Supplementary Volume* XXXII, pp. 143–156.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., & Weiner, A.M. (1987), *Molecular Biology of the Gene*, 4th ed. Menlo Park, CA: Benjamin-Cummings.

Wittgenstein, L. (1953), *Philosophical Investigations*. Oxford: Basil Blackwell, 1958 (second revised edition). Translation by G.E.M. Anscombe.

Wittgenstein, L. (1922), *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul, 1922.

Yčas, M. (1969), *The Biological Code*. Amsterdam: North-Holland.

Index

A

a posteriori xi, xiii, xiv, xvi-xviii, xx, 121, 130, 132, 141-143, 152, 155, 158, 164, 201, 202, 207, 209-213, 216, 217, 219, 220, 222, 231, 233-235, 238, 240, 241, 243, 244, 246
a priori xi-xiv, xvi-xx, 81, 121, 130-137, 141-145, 147-162, 164, 195-213, 216-220, 222, 230-234, 236, 238, 240, 241, 243-246
abstract 61, 62, 66, 140, 165, 209, 220, 221, 226, 230, 237
accessory propositions 158, 159
Ackermann 156, 247
adaptation 217
analysis xvii, xviii, 8, 37, 46, 49, 52, 131, 134, 137, 138, 140, 141, 143, 148, 149, 151-153, 158, 168, 169, 171, 179, 189-191, 202, 204-210, 216, 221, 222, 243, 245, 266
analytic evaluation 143
Analytic Philosophy xi, xv-xviii, 1, 4, 6-8, 21, 25, 39, 40, 52, 54, 84, 98, 100, 105, 125, 127, 163, 164, 198, 205, 218, 225, 226, 228, 229, 237, 240, 242-244
Anderson 156, 247
Anselm xviii, 197
Aplysia 170, 172, 173, 175, 176, 179, 247, 252, 256
Aquinas xviii, 197
Aristotelian 150, 164, 251, 256, 257
Aristotle xviii, 85, 197
associationism 191-193, 213, 216
associative processes xx, 168, 187, 216
asymmetric dependence 109-112, 114-117, 119
atomic propositions 144, 146-148, 150, 164, 203, 206, 230
Augustine xviii, 197
axiomatic xviii, 42, 45-47, 62, 125, 140, 154, 163, 199, 246
æsthetics 222

B

Barcan formula 157, 158, 161
basic ideas 127, 129, 130, 133-136, 145, 164
begriffsschrift 225-229, 242
behavioural holism 16, 17, 21, 67, 69, 77, 89, 93, 95, 96, 98, 104, 105, 107, 109, 116, 121, 123, 180, 191, 229, 244
behaviourism xviii, 16, 17, 21, 39, 44, 59, 60, 69, 77, 87-93, 96-99, 105, 107, 109, 116, 121, 123, 180, 191
Belnap 156, 247
Berkeley 53, 243, 250
Block 109, 111, 112
Boltzmann machine xx
broad meaning 10, 11
Burge 9, 10, 108, 248

C

Cajal 182, 189-193, 213, 248
Carnap xv, 39, 41, 52-56, 105, 248
Cartesian 95, 135, 151, 157
categorematic 5, 6
causation 3, 9, 13, 197
certainty 46, 47, 81, 126, 130, 131, 139-141, 143, 148, 149, 159, 165, 210-213, 222, 241, 245
classical computational architecture xi, xii, 4, 169, 181, 182, 186

Classical Theory of Mind xv, 1, 22, 45, 71, 125-127, 159, 167, 179, 180, 195, 212, 230
cognitive architecture 4, 169, 170, 181, 182, 186, 187
cognitive operations 130, 131, 181, 196, 217, 245
cognitive processes xi, 169, 185, 186, 189, 195, 196, 201, 203, 213
cognitive science 10, 13, 16, 249
cognitive states 175, 179, 185, 187, 190-192
coherence theory of truth 100-104
collateral information 90-93, 96, 97
common-sense 1, 4, 63, 64, 80-83, 127, 168, 179, 182, 189
complex ideas xvi, xviii, 34, 70, 72-75, 99, 123, 128-133, 143, 144, 147, 152, 155, 157, 160, 164, 167, 196, 199, 200, 202, 205, 214-220, 231, 234-236, 241, 244, 245
computational xi, xii, 3, 4, 8-10, 12, 15-17, 21, 44, 72, 107-109, 112, 116, 117, 120, 122, 123, 168-170, 181, 182, 186, 216
computationalism xix, 121, 193
computer analogy 3, 4, 14
concept-acquisition 68, 187
conceptual scheme xiii, xv, 61-66, 68, 101-104
concrete 40, 220, 221, 230, 233, 234, 237
conditional xi, xviii, 7, 8, 51, 87, 126, 131, 134, 152-155, 158, 160-162, 203, 204, 240
confirmation 2, 47, 54-56, 78-81, 83, 84, 86, 100, 159, 199, 201, 209, 210, 216, 245
connectionism xix, 168, 169, 180, 187, 189-191, 193, 216, 217
consciousness xx, 24, 72, 73, 83, 133, 165, 167, 197, 200, 244, 245
Conservative xi, xii, xv, xix, 1, 3, 5, 7, 9, 10, 16, 17, 21, 26, 29, 31, 35, 40, 41, 44, 59, 60, 63, 73, 79, 80, 84, 93, 98, 100, 107, 123, 125, 218, 227, 231
contingency 126, 147, 156, 207, 209, 210, 241
contingent implication 152, 155, 158, 159
contradiction 32, 139, 140, 150, 156, 164, 225, 226
contribution of environment 25, 27, 28, 30, 32, 38, 100, 104, 105
convention xiv, xviii, 22, 36, 87, 101, 105, 106, 120, 125, 222, 229, 246

D

de re 36-38, 123
deducibility xviii, 125, 156, 157
deep-layer 183-186, 188, 195, 196, 203, 213-215
definable 10, 18, 70, 73, 74, 128, 134, 225
definite descriptions 6, 48, 49
definition 13, 27, 31, 134, 148, 200, 201
demonstrative 126, 130, 131, 137, 138, 140, 141, 143, 198, 204, 222, 240, 243
denotation 23, 24, 45, 49, 71, 74, 132, 133, 199-201, 220, 232, 235, 236
denotational theory 65
denoting phrases 48-51, 70
Descartes xviii, 22, 41, 44, 66, 94, 123, 126, 127, 132, 148, 149, 151, 168, 180, 185, 193, 197, 209, 243, 244, 249
determination of extension 12, 27, 28, 30, 36, 40, 60, 63, 64, 70, 74, 122
divided line 220

division of semantic labour xiv, 25, 27, 28, 30, 32, 34-36, 100, 118
Dretske 16
dualism xx, 245

E

e pluribus unum 62-64, 68
education 20, 215, 217, 220
emergent xiv, 31, 173, 235
emotive xvi, 74, 130, 131, 167, 168, 185, 192, 196, 218, 219, 234, 241
empirical ideas xvi, 128, 130, 131, 133, 134, 136, 143, 144, 150, 151, 159, 164, 196-202, 206, 210, 211, 214, 216, 218, 245
empiricism xiii, xiv, 44, 53, 59, 67, 77, 128, 131, 151, 243
entailment xii, 156, 247, 256
epistemic evaluation 42, 78, 126, 135-138, 143-145, 158, 164, 165
epistemic holism xii-xv, 8, 57, 85
epistemic property 133, 135, 147, 149, 150, 155, 157-159, 161, 203, 208, 211, 212
epistemic value 144, 146, 148-150, 164, 199, 202, 203, 206-209, 211, 212
epistemology xv, xvi, 55, 81, 137, 163, 164, 196, 201, 229, 238, 244, 245
error 17, 20, 52, 83, 109, 112-114, 117, 222, 244
essence 22, 33, 34, 37, 40, 44, 45, 72, 75, 83, 86, 97, 167, 200, 228
ethics 140, 222
evaluation xvi, xvii, 42, 78, 81, 82, 84, 86, 126, 130, 131, 135-138, 141-148, 152, 155, 158, 161-165, 199, 203, 205-208, 210-212, 238
ex terminis 46, 141-143, 147, 149, 152-155, 157-159, 161-163, 203, 204, 208, 210, 211
experience xi, xiii, 62, 78, 80, 83, 85, 100, 125, 170, 180, 187, 188, 192, 214, 215, 221, 228
explanation 1, 67, 72, 115, 189, 228, 243
explanatory power 61, 67-69
extension-determiner 10, 31, 41, 59, 60, 63, 64, 70, 79, 98
extension-meaning supervenience xi-xiv, xix, 21, 26, 29, 34-36, 38, 44, 73, 74, 84, 98, 107, 108, 123, 219, 227-229, 231, 233, 234, 236-240
extensional theory xi, 25-27, 38-40, 51, 60, 61, 63-67, 69, 71, 77, 94, 105
externalist 10, 103-105

F

Flexner 192, 249
Fodor xv, 1-22, 44, 72, 75, 107-118, 121, 122, 168, 180, 182, 249, 250
formal model 126-131, 195-197, 203, 205, 206, 230, 231
formalism 43
foundationalism 199
foundations 16, 87, 130, 140, 225, 228, 231, 236, 249, 250, 255
Frege xv, 39, 41-44, 46-50, 52, 55, 60, 72, 77, 87, 105, 164, 165, 225-229, 237, 238, 240, 250, 251
functional role 7, 14

G

Geach 154, 156, 250, 251
gene expression 175, 185, 187
generative mechanism xvi, 128, 129, 131, 145, 180, 185, 195, 196, 214, 231
generative operations xix, 128, 130, 132, 143, 147, 157, 164, 196, 217, 241
genetic code 169, 179, 181
genetic theory of mind 181, 187

geometry 140, 209, 222
GTM 181, 187-193

H

holistic synthesis xvii, 202, 210, 212
Hume 53, 77, 185, 190, 193, 243, 251

I

idea-laden 128, 129, 131, 143, 144, 196
ideational 105, 131, 137, 140, 148, 149, 202, 205, 215, 216, 246
image 18, 41-43, 190
imagistic 41-44, 72
implementation xv, xvi, xviii-xx, 3, 165, 167, 169, 170, 175, 180, 181, 192, 193, 213, 234, 245, 246
implication xviii, 8, 152, 155-162, 164, 201, 203, 242
in pluribus unum 62-64, 68, 77
inference 7, 8, 56, 62, 80, 83, 84, 144, 145, 160-162, 164, 199, 201, 204, 206
information 36, 90-93, 96, 97, 113, 169, 176, 181, 186, 188, 191, 214
innate xi, 13, 18, 63, 75, 117, 126, 138, 139, 141, 151
inscrutability 92, 93, 96, 97
intelligible 95, 220-222, 240
internal realism 59
interpretation 7, 10, 12, 31, 42, 48, 93, 94, 101, 107, 158, 227
introspection 243, 245, 246
introspective 99, 179, 180, 219, 234, 244, 245

K

Kant xviii, xx, 25, 34, 42, 45, 126, 127, 133, 151, 168, 197, 202, 205, 243, 252
knowledge xi, xiii, xiv, xvi, xvii, xx, 27, 28, 30, 33-36, 52, 53, 55, 74, 75, 78, 91, 94, 96, 106, 118, 119, 122, 123, 125, 126, 130, 131, 137-141, 143, 148, 151, 158-160, 170, 195, 197, 199, 201, 202, 210-212, 217, 219, 220, 222, 243, 245, 248, 255, 256, 266

L

language xv, xix, 2-4, 9, 12, 14, 28-31, 43, 45, 46, 50, 53, 62, 65, 68, 69, 75, 85, 87-90, 93-95, 97, 99-101, 103, 117, 120, 121, 132, 153-155, 180, 217-219, 227, 248, 249, 254, 255
Lashley 193, 252
laws 1, 2, 4, 15, 17-19, 54, 56, 78, 81, 85, 113, 158, 159, 216, 225, 228, 250
learning xx, 62, 68, 88-90, 93-95, 116, 117, 121, 122, 168-171, 173-177, 179, 185, 187, 188, 190-192, 210, 215, 247, 249, 252, 256
Leibniz 15, 141, 252
lexical 2, 12, 13, 18, 44, 75, 108, 117, 122, 141
Locke xviii, 22, 33, 41, 44, 45, 66, 71, 74, 77, 93, 123, 126-128, 133, 137-141, 143, 148, 150, 151, 154, 159, 163, 168, 180, 185, 193, 197, 198, 213, 243, 244, 252
logic 4, 8, 11, 43, 45, 78, 86, 87, 127, 128, 130, 140, 145, 150, 152, 154, 156, 158, 161-165, 197-199, 222, 225, 229, 237, 240, 246, 247, 249, 255, 256
logical falsehood 148, 152, 153
logical truth 86, 139, 142, 153, 154, 158, 164
long-term store 186, 196, 213-215, 221

M

Marsh 52, 255, 256
materialism xii, xiii, xx, 53, 167, 234, 245
mathematics 78, 140, 154, 222, 225, 229, 246
meaning as extension-determiner 10, 59, 60, 63, 70, 79
meaning holism 35, 36, 39, 100
meaning-syntax supervenience 5, 120
meaningfulness 16, 40, 46, 47, 55, 59, 60, 63, 64, 70, 101, 227, 233, 240
memory xx, 168-171, 173-176, 179, 182, 185-188, 190-192, 196, 213-215, 217, 221, 246, 247, 249-251, 256
mental code xi, xiv, 4-7, 123, 127, 129, 131-133, 153, 154, 164, 168, 169, 179, 181, 182, 203, 204, 207, 244, 246
mental processes xviii, 2, 3, 8, 72, 182, 201, 204
mental sentence xvi, 132, 135, 167, 180, 198, 211, 220, 241
mental states xii-xiv, xix, 2, 4, 9, 10, 15, 16, 107, 108, 116, 121, 168, 169, 189, 192
mentalism xix, 1, 4, 5, 8, 16, 21, 26, 27, 41-44, 70, 73, 74, 95, 98, 99, 105, 107-109, 112, 116, 117, 121-123, 125, 165, 168, 218, 229, 242
metalogical 137, 141, 142, 152, 160
metaphysical xx, 23, 53, 65, 66, 103, 105, 123, 159, 212, 213, 222, 232-236, 239, 243
metaphysics 40, 41, 60, 61, 63, 65, 69-72, 123, 222, 229, 248, 254
middle-layer 183-185, 188, 195, 215
Mill 42, 87
mind-dependent 66, 95, 201
mind-independent 39, 40, 59, 60, 66, 69, 70, 77, 102, 133, 165, 244
misrepresentation 17, 20, 21, 109, 110, 117
modality xiii, xviii, 8, 11, 125, 126, 131, 147, 148, 155, 157, 162-164, 198, 202, 203, 206, 207, 209, 241
model code 134, 137, 143, 145
Moore 52

N

narrow meaning 10
natural environment 1, 5, 25, 32, 37, 38, 212, 217, 218
natural science xviii, xx, 33, 35, 67, 80, 81, 86, 167, 222, 243, 246
necessary implication 152, 155, 156, 161, 162, 203
necessary truth 7, 8, 125, 207, 241
necessity 8, 126, 147, 148, 162, 207, 247
nomenclature 17, 22, 24, 34, 40, 41, 67, 120, 141
nomic theory of reference 6, 8, 11-13, 16-18, 20, 21, 107-109, 111, 112, 116-120, 122
nominal essence 33, 34, 44, 72, 75, 83, 97, 167, 200
nominalism xii, xiii, 61, 62, 66-69, 72, 98
nominalistic 61-69, 72, 219, 220
nomological xii, xv, 4, 8, 11, 33, 86, 111, 211
non-empirical 42, 81, 121, 128-133, 143, 144, 151, 159, 164, 196, 197, 200, 243, 245
non-existent 47-51, 59, 64
normativity 152, 160
noumenal xvi, 34, 71, 74, 121, 132, 133, 136, 199, 200, 204, 220, 230-232
number terms 46
number-referring 42, 43, 45, 46

O

observation xii, xiii, 18, 80, 81, 106, 138, 216, 222
observational certainty 210, 211
observational synthesis xvii, 202, 210, 211
Ockham xviii, 197

ontic 220
ontological xii, xiii, 16, 40, 61, 67, 71, 72, 104, 180, 181, 255
ontology 41, 61, 65, 67, 68, 72, 103, 167, 187, 195
overt use xiv, 31, 88, 89, 94, 99, 101

P

paraconsistent 226, 227, 229, 237
particulars 1, 5, 6, 8, 23, 24, 38, 40, 44, 45, 51, 52, 63, 65, 66, 71, 98, 164, 198, 226, 230, 233, 234
phenomenalism 53, 80
Plato xviii, xx, 40, 41, 197, 202, 209, 242, 244
Plato's beard 39-41, 59-61, 63, 69, 70, 94, 97, 105, 244
possibility 8, 11, 45, 57, 83, 126, 147, 158, 159, 169, 207
possible worlds xii, xiii, xv, xviii, 7, 8, 15, 16, 20, 21, 36, 111, 125, 142, 147, 156, 157, 198
pragmatic xii, 29, 31, 53-55, 59, 61, 67, 79, 95, 96, 100, 105, 159, 210, 212, 213, 222, 246
propositional attitudes 1, 4, 8, 9
psychic cell xvi, 182, 183, 188, 190, 195, 196, 199, 201, 203, 213-216
psychic-cell architecture 184
psychophysical 17-20, 108, 109, 117, 118, 120, 200
public language xv, 2, 12, 14, 45, 75, 99, 120, 155, 217-219
public symbols xii, xiv, 16, 25, 73, 105, 109, 121, 195, 219
Putnam xv, 9, 10, 12, 22, 25-36, 39, 44, 60, 93, 100-105, 108, 114, 117, 119, 122, 254

Q

quadripartite division 186
Quine xv, 22, 27, 31, 35, 39-41, 44, 57, 59-74, 77-81, 84-93, 95-101, 105, 125, 162, 163, 229, 244, 254, 255

R

Radical xi-xiii, xv, xviii, xix, 17, 59, 77, 90, 91, 93, 98, 244
rational processes xiv, 182, 185, 187-189, 196, 214, 216
rationalism xi, xiii, 1, 13, 16, 17, 75, 107, 151
real essence 22, 33, 34, 44, 83, 86, 200
real property 23, 24, 33, 230-232, 236
realism 29, 31, 59, 63, 65, 67, 100, 104, 105
realistic 61-66, 68, 77
recall 48, 86, 100, 131, 158, 175, 182, 183, 186, 187, 232
recollection xx, 214, 215, 246
reference xi-xv, 6-8, 10-13, 16-21, 23, 24, 29-31, 36, 40, 45, 49, 52, 59, 60, 62-64, 67, 70, 74, 75, 92, 93, 96-98, 100, 101, 103-105, 107-109, 111, 112, 116-123, 142, 147, 169, 220, 227, 232, 235, 236
referent 24, 30, 37, 40, 44, 47, 50, 52, 225, 227, 229, 233, 237-240
referential theory of meaning xi, xii, 5, 6, 15, 16, 107-109, 111, 112, 120-122
reflection 74, 127, 130, 157
regimentation 63, 95, 217-220, 235, 236
relativism xv, 199
relativity 11, 65, 67, 101-104, 255
representation 2, 6, 12, 20, 22-24, 41, 43, 45, 67, 74, 84, 104, 109, 112, 117, 118, 120, 121, 126, 132, 133, 168, 169, 181, 183, 188, 189, 195-197, 199, 201, 254
retention 170, 173, 174, 176, 178, 179, 187, 213-215
revisability 79-81, 83, 85, 86
Richards 52, 255
Russell xv, 39, 41, 47-53, 55, 87, 105, 164, 225, 228, 229, 237, 238, 240, 255, 256
Russell's paradox xv, 195, 223, 225, 229

S

satisfaction theory of reference 100, 103-105
Searle 119, 180, 256
semantic holism 7, 32, 39-41, 54, 56, 57, 59, 77, 80, 83-85, 93, 94, 96, 100-103, 105
semantic identity xiv, 2, 4, 5, 16, 37, 44, 86, 93, 94, 97, 107, 109, 130, 131, 136, 138, 139, 142, 143, 146, 148, 149, 151-155, 160, 161, 164, 202-206, 208, 216, 219, 230, 233, 241
semantic property 1, 5, 31, 72, 200
sensation 53, 72, 74, 127, 130, 200
sensitisation 170-175, 253
sensorium 19, 118, 185, 186, 196, 214, 215, 221
sentence-based xvi, xviii, 8, 39, 45, 47, 48, 50-52, 54, 55
set theory 43, 225-229, 231, 232, 236, 240-242
set-membership 43, 45, 230
simple ideas 34, 70, 71, 73, 74, 99, 127-129, 132, 133, 139, 141, 144, 146, 148-151, 153, 154, 160, 164, 167, 196, 199-201, 210, 218, 219, 221, 231, 241, 244
Smiley 154, 156, 157, 256
sociological phase 18, 19, 109, 116-118, 121-123
state of affairs xvi, 23, 24, 83, 84, 132, 136, 145, 149, 150, 167, 188, 198, 199, 201, 215, 230
statement xi, xii, xiv, 41, 43, 45-47, 49, 52, 55, 56, 77-86, 102, 220, 240
stereotype 27-29, 35
stimulus-dependence 106
stimulus-independence 95
Strawson 154, 156, 256
supervenience xi-xiv, xix, 5, 8-10, 12, 13, 21, 22, 26, 29, 34-36, 38, 44, 45, 73, 74, 84, 98, 107, 108, 120, 123, 218, 219, 227-229, 231-234, 236-242
surface-layer 184, 188, 195, 215
symbolic system xviii, xix, 2, 7, 11, 16, 102, 104, 136, 142, 167, 195, 196, 199, 242
syncategorematic 5-7
synonymy 29, 91, 92
syntax 3-5, 13-15, 28, 72, 108, 120, 163, 167, 168, 196, 205, 244, 245
synthesis xvii, xviii, 143, 149, 158, 174, 177-179, 191, 192, 202, 204-212, 214, 216, 221, 222, 243, 245, 251, 253, 256, 266

T

term-based xv, xvi, 8, 10, 16, 21, 40-48, 50, 52, 55, 59, 64-67, 73, 105, 123, 132, 134, 141, 142, 150, 199, 204, 218, 231
theology 140, 222
transfer 183, 187, 191, 214, 252, 253
translation 31, 89-93, 95-98, 100, 101, 179, 250, 252, 255, 257
trifling propositions 149, 151, 154, 203, 205, 208
truth-condition 5-7, 9, 12, 21-24, 78-80, 84, 93, 98, 230, 233, 235, 238-241
truth-conditional theory 7, 8, 51, 87, 240
truth-functions 128, 206
Turing machine xx
Twin-Earth 9-12, 27, 31, 105, 107, 114, 249

U

undefinability 73, 122
uniformity 34, 50, 68, 219, 220
universals 6, 22, 23, 39, 40, 59-61, 65-72, 77, 132, 136, 164, 199, 235

V

validity xii, xviii, 126, 161, 162, 229, 242, 246
verbal use xiv, 38, 94, 95

verificationism 33, 39, 47, 52, 53, 55, 59, 60, 77, 82-84, 87, 93, 97, 105
vertebrates 175, 248
von Neumann xi, xiv, 4, 169
Von Wright 154, 156, 257

W

Watling 154, 156, 257
Wittgenstein xviii, xix, 53, 59, 64, 70, 229, 244, 257
working memory 176, 186, 188, 196, 214, 215, 217, 221

STUDIES IN COGNITIVE SYSTEMS

1. J.H. Fetzer (ed.): *Aspects of Artificial Intelligence.* 1988
   ISBN 1-55608-037-9; Pb 1-55608-038-7
2. J. Kulas, J.H. Fetzer and T.L. Rankin (eds.): *Philosophy, Language, and Artificial Intelligence.* Resources for Processing Natural Language. 1988 ISBN 1-55608-073-5
3. D.J. Cole, J.H. Fetzer and T.L. Rankin (eds.): *Philosophy, Mind and Cognitive Inquiry.* Resources for Understanding Mental Processes. 1990 ISBN 0-7923-0427-6
4. J.H. Fetzer: *Artificial Intelligence: Its Scope and Limits.* 1990
   ISBN 0-7923-0505-1; Pb 0-7923-0548-5
5. H.E. Kyburg, Jr., R.P. Loui and G.N. Carlson (eds.): *Knowledge Representation and Defeasible Reasoning.* 1990 ISBN 0-7923-0677-5
6. J.H. Fetzer (ed.): *Epistemology and Cognition.* 1991 ISBN 0-7923-0892-1
7. E.C. Way: *Knowledge Representation and Metaphor.* 1991 ISBN 0-7923-1005-5
8. J. Dinsmore: *Partitioned Representations.* A Study in Mental Representation, Language Understanding and Linguistic Structure. 1991 ISBN 0-7923-1348-8
9. T. Horgan and J. Tienson (eds.): *Connectionism and the Philosophy of Mind.* 1991
   ISBN 0-7923-1482-4
10. J.A. Michon and A. Akyürek (eds.): *Soar: A Cognitive Architecture in Perspective.* 1992
    ISBN 0-7923-1660-6
11. S.C. Coval and P.G. Campbell: *Agency in Action.* The Practical Rational Agency Machine. 1992 ISBN 0-7923-1661-4
12. S. Bringsjord: *What Robots Can and Can't Be.* 1992 ISBN 0-7923-1662-2
13. B. Indurkhya: *Metaphor and Cognition.* An Interactionist Approach. 1992
    ISBN 0-7923-1687-8
14. T.R. Colburn, J.H. Fetzer and T.L. Rankin (eds.): *Program Verification.* Fundamental Issues in Computer Science. 1993 ISBN 0-7923-1965-6
15. M. Kamppinen (ed.): *Consciousness, Cognitive Schemata, and Relativism.* Multidisciplinary Explorations in Cognitive Science. 1993 ISBN 0-7923-2275-4
16. T.L. Smith: *Behavior and its Causes.* Philosophical Foundations of Operant Psychology. 1994
    ISBN 0-7923-2815-9
17. T. Dartnall (ed.): *Artificial Intelligence and Creativity.* An Interdisciplinary Approach. 1994
    ISBN 0-7923-3061-7
18. P. Naur: *Knowing and the Mystique of Logic and Rules.* 1995 ISBN 0-7923-3680-1
19. P. Novak: *Mental Symbols.* A Defence of the Classical Theory of Mind. 1997
    ISBN 0-7923-4370-0

KLUWER ACADEMIC PUBLISHERS – DORDRECHT / BOSTON / LONDON